Preprint typeset in JHEP style - PAPER VERSION UCB-PTH-08/71

Selfduality and Chern–Simons Theory

Ori J. Ganor and Yoon Pyo Hong

Department of Physics, University of California, Berkeley, CA 94720
Emails: `origa@socrates.berkeley.edu,yph@berkeley.edu`

Abstract: We propose a relation between the operator of S-duality (of $\mathcal{N}=4$ super Yang–Mills theory in 3+1D) and a topological theory in one dimension lower. We construct the topological theory by compactifying $\mathcal{N}=4$ super Yang–Mills on S^1 with an S-duality and R-symmetry twist. The S-duality twist requires a selfdual coupling constant. We argue that for a sufficiently low rank of the gauge group the three-dimensional low-energy description is a topological theory, which we conjecture to be a pure Chern–Simons theory. This conjecture implies a connection between the action of mirror symmetry on the sigma-model with Hitchin's moduli space as target space and geometric quantization of the moduli space of flat connections on a Riemann surface.

Keywords: S-Duality, Chern–Simons, Supersymmetry, Flat Connections, Hitchin Space, T-duality, Mirror symmetry, Geometric quantization.

Contents

---

1. Introduction

In the last 14 years there has been a lot of progress on the subject of S-duality of $\mathcal{N} = 4$ super Yang–Mills theory (SYM). The conjecture [1, 2, 3] has passed many elaborate tests, including those in the following list: the number of BPS dyons obeys S-duality [4], the partition function of a topologically twisted theory is S-dual [5], and after a supersymmetric compactification on a Riemann surface S-duality reduces to mirror symmetry in the low-energy limit [6, 7]. The moduli space of mass-deformed $\mathcal{N} = 4$ has been established to obey S-duality [8], the action of S-duality on operators has been deduced for many local operators [9], as well as Wilson loops and 't Hooft loops [10], it was demonstrated that low-energy states of the toroidally compactified theory also obey S-duality [11], and a comprehensive framework for determining the action of S-duality on BPS boundary conditions [12] has been developed in [13]. S-duality also fits nicely within the framework of the AdS/CFT correspondence [14].

New insights on S-duality emerged from various sources: Witten's conjecture [15] that $\mathcal{N} = 4$ SYM is the low-energy limit of a T^2 compactification of a six-dimensional conformal field theory, the $(2, 0)$-theory, gave rise to a geometrical realization of S-duality with simply-laced groups, and this has been generalized to other groups as well [16]; M(atrix) theory [17] and its application to the $(2, 0)$-theory [18, 19] therefore led to a conjectured S-duality invariant formulation of $\mathcal{N} = 4$ SYM [20]. More recently, new insight has emerged about the connection between S-duality and the geometric Langlands program [21].

But S-duality still remains a mystery. What is required is an operator $\mathcal{S}$ that transforms a state in the Hilbert space of $\mathcal{N} = 4$ SYM to a dual state in the Hilbert space of the dual theory. (The approach of studying the Hilbert space of a gauge theory was a fruitful one in three-dimensions [22, 23] and also an interesting direction in four-dimensions [24].) The operator $\mathcal{S}$ is directly related to the action of S-duality on

boundary conditions, as defined in [13].[1] In this paper we will attempt to gather new clues about the nature of the operator $\mathcal{S}$. We will argue that this operator is related to a three-dimensional nonlocal topological field theory. We will also argue that under certain restrictions a *local* topological field theory emerges. The physical questions that define observables in this local theory are as follows. Suppose we compactify $\mathcal{N} = 4$ SYM on S^1 (parameterized by a periodic coordinate $0 \le x_3 < 2\pi R$), which we shall treat as (Euclidean) time, but instead of setting up periodic boundary conditions whereby the quantum states of the system at $x_3 = 0$ and $x_3 = 2\pi R$ are required to agree, we instead require that the state at $x_3 = 2\pi R$ is the *S-dual* of the state at $x_3 = 0$. This is an "S-duality twist," which is possible at certain special selfdual values of the coupling constant. Our question now is: *what is the low-energy effective three-dimensional description of this theory?* With a few additional modifications and restrictions, we will argue that this is a topological theory, and we propose that it is a Chern–Simons theory. Observables can then be constructed from Wilson lines that are at a constant x_3 and we expect them to reduce to Wilson lines in Chern–Simons theory. The expectation values of these observables are related to knot invariants [25]. [For the rest of this paper, unless otherwise noted, x_3 will be understood as a spatial (as opposed temporal) direction.]

The paper is organized as follows. In §2 we construct the circle compactification with the S-duality twist. We also introduce an R-symmetry twist in order to preserve supersymmetry and reduce the number of zero modes, and we add restrictions on the rank of the gauge group in order to eliminate all zero modes. In §3 we study the S-duality operator from the Hilbert space perspective, and argue that a topological three-dimensional action can be constructed from it. In §4 we take a detour to study a related problem in two-dimensional conformal field theory. We discuss a σ-model with target space T^d at a selfdual point in moduli space, and show that compactification on S^1 with a T-duality twist reduces in low-energy to a 0+1D topological theory which can be identified with geometric quantization of the target space. We then discuss a possible generalization of this result to $\mathcal{N} = (2, 2)$ supersymmetric σ-models which are selfdual under mirror symmetry. In §5 we return to four-dimensional $\mathcal{N} = 4$ SYM and describe the abelian case where exact results are well-known. In §6 we turn to the S-duality and R-symmetry twisted circle compactification of nonabelian $\mathcal{N} = 4$ SYM. We propose that the low-energy topological theory is pure Chern–Simons theory, and outline a test of this conjecture whereby we compactify on a Riemann surface of genus 2 and count the number of vacua. Because of several unknown (at least to us) signs in the action of S-duality, we only get partial results. We conclude in §7 with a discussion

[1] We are grateful to E. Witten for explaining this point to us.

and suggestion for further explorations.

2. The problem

Four-dimensional $\mathcal{N} = 4$ super Yang–Mills theory with gauge group $U(n)$ is believed to possess $\mathrm{SL}(2,\mathbb{Z})$-duality. The complex coupling constant

$$\tau \equiv \frac{4\pi i}{g_{\mathrm{YM}}^2} + \frac{\theta}{2\pi}$$

transforms under an element

$$\begin{pmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$$

as

$$\tau \to \frac{\mathbf{a}\tau + \mathbf{b}}{\mathbf{c}\tau + \mathbf{d}} .$$

In this paper we are concentrating on *selfduality*, which occurs at values of τ for which there exists an element $\mathbf{s} \in \mathrm{SL}(2,\mathbb{Z})$, other than $\mathbf{I} \equiv \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ or $-\mathbf{I}$, that leaves τ invariant. A selfdual theory can be compactified on an S^1 with an $\mathbf{s}$-twist. We are interested in the 2+1D low-energy limit of this setting. We will add a few more ingredients and restrictions in order to preserve supersymmetry and to eliminate zero modes, and we will argue that the resulting 2+1D theory is a *topological field theory.* The problem is: *what is that topological field theory?* We now turn to the details.

2.1 S-duality twist

Up to $\mathrm{SL}(2,\mathbb{Z})$-conjugation, selfduality occurs in the following two cases:

1. $\tau = i$ is fixed by the $\mathbb{Z}_4$ subgroup of $\mathrm{SL}(2,\mathbb{Z})$ generated by

$$\mathbf{s}' \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}); \tag{2.1}$$

2. $\tau = e^{\pi i/3}$ is fixed by the $\mathbb{Z}_6$ subgroup generated by

$$\mathbf{s}'' \equiv \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z}) . \tag{2.2}$$

For a selfdual value τ, any duality transformation $\mathbf{s}$ that fixes it is a symmetry. The elements $\mathbf{s}'$ or $\mathbf{s}''$ above generate subgroups of $\mathrm{SL}(2,\mathbb{Z})$ (either $\mathbb{Z}_4$ or $\mathbb{Z}_6$) which are

discrete gauge symmetries. For a given τ, we will denote the subgroup of $\mathrm{SL}(2,\mathbb{Z})$ that fixes it by $\mathbb{S}_\tau$. On a manifold X with nontrivial first homotopy group $\pi_1(X)$, we can formulate the theory with boundary conditions that are twisted by elements of $\mathbb{S}_\tau$ along nontrivial loops. The complete set of choices for such boundary conditions is given by the set of homomorphisms (maps that preserve the group structure) from $\pi_1(X)$ to $\mathbb{S}_\tau$. In this paper, we will only study the case $X = S^1 \times M_3$, where M_3 is some 3-manifold, with an $\mathrm{SL}(2,\mathbb{Z})$-twist only along S^1.

Ignoring M_3 for the moment, we have 3+1D $\mathcal{N}=4$ $U(n)$ SYM compactified on S^1 with an $\mathbf{s}$-twist. We pick $\mathbf{s}$ and τ from the following list of choices:

1. $\tau = i$ and $\mathbf{s} = \mathbf{s}' \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$;

2. $\tau = e^{\pi i/3}$ and either $\mathbf{s} = \mathbf{s}'' \equiv \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$ or $\mathbf{s} = -\mathbf{s}'' = \mathbf{s}''^4 \equiv \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix} \in \mathrm{SL}(2,\mathbb{Z})$;

All other possible $\mathbf{s}$'s are $\mathrm{SL}(2,\mathbb{Z})$-conjugate to those in the list, or their inverses (which give theories that are physically equivalent after a parity transformation).

Why should we believe that an $\mathrm{SL}(2,\mathbb{Z})$-twist is allowed? Duality symmetries have been used quite extensively in string theory to twist boundary conditions. (See for instance [26]-[30] for some old and some recent examples.) Moreover, The Euclidean partition function on $S^1 \times M_3$ with the $\mathbf{s}$-twist is easy to define — we simply treat S^1 with radius R as the (Euclidean) time direction and calculate $\mathrm{tr}\{(-1)^F e^{-2\pi R H}\hat{\mathcal{S}}\}$, where F is the fermion number, H is the Hamiltonian on M_3, and $\hat{\mathcal{S}}$ is the operator that corresponds to the action of $\mathbf{s}$ on the Hilbert space. (Here, we momentarily think of x_3 as temporal, but from now on, until otherwise stated, we return to thinking about it as a spatial direction.) Moreover, the $\mathbf{s}$-twist can be given a purely geometrical construction in terms of the 5+1D $(2,0)$-theory. As Witten's conjecture goes [15], 3+1D $\mathcal{N}=4$ $U(n)$ SYM with coupling constant τ can be realized by compactifying the 5+1D $(2,0)$ theory (the low-energy limit of n coincident M5-branes [31]) on a T^2 with complex structure τ and area $\mathcal{A}$ and taking the $\mathcal{A} \to 0$ limit. The T^2 can be described as the quotient $\mathbb{C}/(\mathbb{Z}+\mathbb{Z}\tau)$ (where $\mathbb{C}$ is the complex plane and $\mathbb{Z}+\mathbb{Z}\tau$ is the lattice generated by 1 and τ), and if $z \sim z+1 \sim z+\tau$ is a complex coordinate on $\mathbb{C}$, then the $\mathrm{SL}(2,\mathbb{Z})$ duality transformation $\mathbf{s} = \begin{pmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{pmatrix}$ is realized as a change of basis $1 \mapsto \mathbf{c}\tau + \mathbf{d}$ and $\tau \mapsto \mathbf{a}\tau + \mathbf{b}$. If τ is fixed by $\mathbf{s}$ then $\mathbf{a}\tau + \mathbf{b} = (\mathbf{c}\tau + \mathbf{d})\tau$ and it is then easy to check that, assuming

$\operatorname{Im}\tau > 0$, $|\mathbf{c}\tau+\mathbf{d}| = 1$, so we can write

$$\mathbf{c}\tau + \mathbf{d} = e^{i\upsilon}\,, \tag{2.3}$$

for some phase υ. For $\tau = i$ and $\mathbf{s} = \mathbf{s}'$ we have $\upsilon = \frac{\pi}{2}$; for $\tau = e^{\pi i/3}$ and $\mathbf{s} = \mathbf{s}''$ we have $\upsilon = \frac{\pi}{3}$ and for $\mathbf{s} = -\mathbf{s}''$ we have $\upsilon = \frac{4\pi}{3}$. The $\mathbf{s}$-duality transformation can then be realized as rotation of the $\mathbb{C}$ plane by an angle υ, which preserves the lattice $\mathbb{Z}+\mathbb{Z}\tau$.

A geometric realization of the $\mathbf{s}$-twisted compactification of $\mathcal{N}=4$ SYM is now clear. We take a circle S^1 of radius R, parameterized by $0 \le x_3 < 2\pi R$ (we reserve x_0, x_1, x_2 for coordinates on M_3), and compactify the $(2,0)$ theory on the space parameterized by (x_3, z) with identifications

$$(x_3, z) \sim (x_3, z+1) \sim (x_3, z+\tau) \sim (x_3 + 2\pi, e^{i\upsilon} z), \tag{2.4}$$

and metric

$$ds^2 = dx_3^2 + \frac{\mathcal{A}}{\operatorname{Im}\tau}|dz|^2\,. \tag{2.5}$$

In the limit $\mathcal{A} \to 0$ we recover the $\mathcal{N}=4$ $U(n)$ SYM theory compactified on S^1 with an $\mathbf{s}$-twist.

As it stands, compactification of $\mathcal{N}=4$ SYM on S^1 with an $\mathbf{s}$-twist breaks all the supersymmetries. The supercharges $Q_{a\alpha}$ ($a = 1,\dots,4$ and $\alpha = 1,2$) are in the representation $\mathbf{2}$ of the Lorentz group $SO(3,1)$ and $\overline{\mathbf{4}}$ of the R-symmetry group $SU(4)_R$, and their complex conjugates $\overline{Q}^a_{\dot\alpha}$ are in the $\mathbf{2}'$ of $SO(3,1)$ and $\mathbf{4}$ of $SU(4)_R$. As explained in [21], the supercharges transform under $\mathbf{s}$ as

$$\mathbf{s} : Q_{a\alpha} \to \left(\frac{\mathbf{c}\tau+\mathbf{d}}{|\mathbf{c}\tau+\mathbf{d}|}\right)^{1/2} Q_{a\alpha} = e^{\frac{i\upsilon}{2}} Q_{a\alpha}\,. \tag{2.6}$$

This is easy to see from the $(2,0)$-theory realization mentioned above—the duality corresponds to rotation by an angle υ on a plane of the T^2. Since no linear combination of the supercharges is invariant under $\mathbf{s}$-duality, no supersymmetry is preserved by the $\mathbf{s}$-twist. We would, however, like to preserve some amount of supersymmetry so as to be able to use Witten-index techniques later on. For this purpose we will now add an R-symmetry twist. As a bonus, we will see that an appropriate twist also eliminates some unwanted zero modes of scalar fields.

2.2 R-symmetry twist

The fields of 3+1D $\mathcal{N}=4$ $U(n)$ SYM are as follows: a gauge field A_μ, 6 adjoint-valued scalar fields Φ^I ($I = 1,\dots,6$), and 4 spinor fields ψ^a_α ($a = 1,\dots,4$ and $\alpha = 1,2$) and their complex conjugates $\overline{\psi}_{a\dot\alpha}$ ($a = 1,\dots,4$ and $\dot\alpha = \dot 1, \dot 2$).

The R-symmetry is $\mathrm{Spin}(6) = SU(4)$. The 6 scalars Φ^I are in the real representation $\mathbf{6}$, the fermions ψ^a_α are in the complex representation $\mathbf{4}$, and $\overline{\psi}_{a\dot{\alpha}}$ are in $\overline{\mathbf{4}}$. We pick bases of $\mathbf{4}$ and $\overline{\mathbf{4}}$ such that a diagonal element

$$\gamma \equiv \begin{pmatrix} e^{i\varphi_1} & & & \\ & e^{i\varphi_2} & & \\ & & e^{i\varphi_3} & \\ & & & e^{i\varphi_4} \end{pmatrix} \in SU(4)_R\,, \qquad \left(\sum_a \varphi_a = 0\right)\,, \tag{2.7}$$

acts as

$$\gamma(\psi^a_\alpha) = e^{i\varphi_a}\psi^a_\alpha\,, \qquad \gamma(\overline{\psi}_{a\alpha}) = e^{-i\varphi_a}\overline{\psi}_{a\alpha}\,.$$

We also pick a basis of $\mathbf{6}$ such that the representation of γ in $SO(6)$ acts on the scalars as

$$\left.\begin{aligned} \gamma(\Phi^{2a-1}) &= \Phi^{2a-1}\cos(\varphi_a+\varphi_4) - \Phi^{2a}\sin(\varphi_a+\varphi_4) \\ \gamma(\Phi^{2a}) &= \Phi^{2a-1}\sin(\varphi_a+\varphi_4) + \Phi^{2a-1}\cos(\varphi_a+\varphi_4) \end{aligned}\right\} \tag{2.8}$$

for $a = 1, 2, 3$. On occasion, we will suppress the indices on $\psi, \overline{\psi}, \Phi$ and denote the action of γ on the fields by ψ^γ, $\overline{\psi}^\gamma$, and Φ^γ.

Let $0 \le x_3 < 2\pi R$ be a periodic coordinate on S^1. We now augment the $\mathbf{s}$-twist from §2.1 by an R-symmetry twist as follows. (See also [32] for a similar use of R-symmetry in a different context.) Without the $\mathbf{s}$-twist, an R-symmetry twist is simply a modification of the boundary conditions for the scalars and spinors of $\mathcal{N} = 4$ SYM to

$$\Phi(2\pi R) = \Phi(0)^\gamma\,, \qquad \psi(2\pi R) = \psi(0)^\gamma\,, \tag{2.9}$$

where γ is some element of the R-symmetry group, and only the x_3 argument is shown in (2.9). This twist is independent of the position of the origin $x_3 = 0$, and we can combine it with the $\mathbf{s}$-twist by modifying the boundary conditions anywhere along S^1.

Let us now discuss the amount of supersymmetry that is preserved. For a generic γ in the form (2.7), no supersymmetry is preserved. Some supersymmetry can be preserved for special choices of the phases φ_a ($a = 1, \ldots, 4$) in (2.7). The combined effect of $\mathbf{s}$ and γ on a supersymmetry generator $Q_{a\alpha}$ is given by

$$Q_{a\alpha} \to e^{\frac{i\upsilon}{2} - i\varphi_a} Q_{a\alpha}\,. \tag{2.10}$$

In general, we get $N = 2r$ supersymmetry in 3D, where r is the number of indices a for which $e^{i\varphi_a} = e^{i\upsilon/2}$, according to (2.10). We thus can get as high as $\mathcal{N} = 6$ in 3D. This maximal amount of supersymmetry arises with

$$\gamma = \begin{pmatrix} e^{\frac{i}{2}\upsilon} & & & \\ & e^{\frac{i}{2}\upsilon} & & \\ & & e^{\frac{i}{2}\upsilon} & \\ & & & e^{-\frac{3i}{2}\upsilon} \end{pmatrix} \in SU(4)_R\,. \tag{2.11}$$

We get $\mathcal{N}=4$ with

$$\gamma = \begin{pmatrix} e^{\frac{i}{2}\upsilon} & & & \\ & e^{\frac{i}{2}\upsilon} & & \\ & & e^{-i(\upsilon+\varphi_4)} & \\ & & & e^{i\varphi_4} \end{pmatrix} \in SU(4)_R \,, \tag{2.12}$$

for any choice of φ_4, and we get $\mathcal{N}=2$ with

$$\gamma = \begin{pmatrix} e^{\frac{i}{2}\upsilon} & & & \\ & e^{-i(\varphi_3+\varphi_4-\frac{1}{2}\upsilon)} & & \\ & & e^{i\varphi_3} & \\ & & & e^{i\varphi_4} \end{pmatrix} \in SU(4)_R \,, \tag{2.13}$$

for any choice of φ_3, φ_4.

To summarize, our setting is $\mathcal{N}=4$ SYM compactified on S^1 with an S-duality twist $\mathbf{s}$ and an R-symmetry twist γ along the circle. We are interested in the limit where the radius of the circle R shrinks to zero. For the rest of this paper, we will take the twist (2.11) which preserves the maximal amount of $\mathcal{N}=6$ supersymmetry.

2.3 Zero modes

Our goal is to analyze the low-energy limit of the compactification of $\mathcal{N}=4$ $U(n)$ SYM on S^1 with both S-duality twist discussed in §2.1 and R-symmetry twist discussed in §2.2. We would like to claim that, for sufficiently small values of n, the low-energy limit leads to a nontrivial 2+1D topological field theory. The restriction on n comes because for large values of n the claim is defeated by the presence of zero modes as we shall now discuss.

We can attempt to construct a "Higgs phase" of the theory as follows. If the radius R of S^1 is sufficiently large (compared to a length scale to be defined shortly), we can first reduce the 3+1D $\mathcal{N}=4$ theory to its Coulomb branch, which at a generic point of the moduli space is described by n free $\mathcal{N}=4$ $U(1)$ vector multiplets. We will denote the scalars in the k-th multiplet by ϕ_k^I ($k=1,\ldots,n$ and $I=1,\ldots,6$) and the electric and magnetic fields in the same multiplet by $\vec{E}_k$ and $\vec{B}_k$, respectively. As in §2.2, we will suppress the R-symmetry index I of the scalars. The next step is to compactify this $U(1)^n$ gauge theory on S^1 with an SL$(2,\mathbb{Z})$-duality and R-symmetry twist. Assuming that

$$\sum_{I=1}^{6} |\langle\Phi_k^I\rangle - \langle\Phi_l^I\rangle|^2 \gg \frac{1}{R^2}\,, \qquad 1\le k<l\le n$$

(where $\langle\Phi^I_k\rangle$ denote the eigenvalues of the VEV of Φ^I), the two-step reduction to low energy is self-consistent.

At first look, the combined $\mathrm{SL}(2,\mathbb{Z})$ and R-symmetry twists set the following low-energy boundary conditions:

$$\vec{E}_k(2\pi R) = \mathbf{a}\vec{E}_k(0) + \mathbf{b}\vec{B}_k(0)\,, \quad \vec{B}_k(2\pi R) = \mathbf{c}\vec{E}_k(0) + \mathbf{d}\vec{B}_k(0)\,,$$
$$\phi_k(2\pi R) = \phi_k(0)^\gamma\,,$$

where $\mathbf{s} \equiv \begin{pmatrix}\mathbf{a} & \mathbf{b}\\ \mathbf{c} & \mathbf{d}\end{pmatrix}$ is the appropriate $\mathrm{SL}(2,\mathbb{Z})$ element from §2.1 and γ is R-symmetry element from (2.11), which acts on the suppressed I index of ϕ^I_k. The first two boundary conditions can also be written as

$$\vec{E}_k(0) - \tau\vec{B}_k(0) = (\mathbf{c}\tau + \mathbf{d})(\vec{E}_k(2\pi R) - \tau\vec{B}_k(2\pi R)) = e^{i\upsilon}(\vec{E}_k(2\pi R) - \tau\vec{B}_k(2\pi R))\,.$$

The zero modes can be found by taking the fields to be independent of the S^1 coordinate, setting $\vec{E}_k(2\pi R) = \vec{E}_k(0)$, $\vec{B}_k(2\pi R) = \vec{B}_k(0)$, and $\phi_k(2\pi R) = \phi_k(0)$. Since none of the $\mathrm{SL}(2,\mathbb{Z})$ twists discussed in §2.1 (neither $\mathbf{s}'$ nor $\pm\mathbf{s}''$) have an eigenvalue 1, none of the phases υ of (2.3) are zero, and there are no zero modes of the vector fields with these boundary conditions. Since γ of (2.11) has no eigenvalue 1, there are also no zero modes of the scalar fields.

However, when the gauge symmetry is broken as $U(n) \to U(1)^n$ by the VEVs $\langle\Phi^I_k\rangle$, we also get an action of the Weyl group $S_n \subset U(n)$ on the low-energy fields. It acts by permuting the indices $k = 1, \dots, n$. Since S_n is a remnant of the gauge group $U(n)$, we are allowed to consider sectors for which the boundary conditions along S^1 are twisted by an element $\sigma \in S_n$. The boundary conditions in this sector become

$$\vec{E}_{\sigma(k)}(0) - \tau\vec{B}_{\sigma(k)}(0) = e^{i\upsilon}(\vec{E}_k(2\pi R) - \tau\vec{B}_k(2\pi R))\,, \qquad \phi_{\sigma(k)}(0)^\gamma = \phi_k(2\pi R)\,,$$

where σ is understood as a permutation map $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$. We get a surviving 2+1D low-energy mode of the vectors for any linear combination $\sum_{k=1}^n C_k(\vec{E}_k - \tau\vec{B}_k)$ such that the coefficients satisfy

$$C_k = e^{i\upsilon}C_{\sigma(k)}\,, \qquad k = 1, \dots, n. \tag{2.14}$$

In other words, $e^{i\upsilon}$ must be an eigenvalue of σ in its fundamental representation.

Similarly, from (2.11) we see that the eigenvalues of γ in the representation $\mathbf{6}$ of $SO(6)$ are $e^{\pm i\upsilon}$ (each one occurring with a multiplicity of 3). Thus, we also get three complex scalar zero modes of the form $\sum_{I=1}^6\sum_{k=1}^n \lambda^I_{(a)}C_k\phi^I_k$. Here $\lambda^I_{(a)}$, for $a = 1, 2, 3$,

are the three eigenvectors of γ with eigenvalue e^{-iv} (which cancels the phase of (2.14) to give rise to the zero modes). In the basis of (2.8), we can take

$$\sum \lambda^I_{(1)}\phi^I = \phi^1 + i\phi^2 \,, \qquad \sum \lambda^I_{(2)}\phi^I = \phi^3 + i\phi^4 \,, \qquad \sum \lambda^I_{(3)}\phi^I = \phi^5 + i\phi^6 \,.$$

for any set of coefficients $\{C_k\}$ that satisfy (2.14).

For $\tau = i$ and $\mathbf{s} = \mathbf{s}'$ we have $v = \frac{\pi}{2}$, and a nonzero solution to (2.14) exists only if the Weyl group has an element of order four, and hence only if $n \geq 4$. For the threshold value $n = 4$, a nonzero solution requires that σ act as $\sigma : (1, 2, 3, 4) \mapsto (2, 3, 4, 1)$, up to conjugation. The space of solutions is then one-dimensional, and the low-energy limit of the compactified $\mathcal{N} = 4$ theory is given by a single 2+1D free $\mathcal{N} = 8$ multiplet of 8 real scalar fields (the three complex scalar zero modes give 6 fields, and the vector field gives rise to 2 scalar fields) with moduli space $(\mathbb{C}^3 \times T^2)/\mathbb{Z}_4$. The $\mathbb{C}^3$ factor corresponds to the three scalar zero modes, and the T^2 (with complex structure $\tau = i$) factor comes out of the vector zero mode. The action of $\mathbb{Z}_4$ corresponds to multiplication by i for each of the four factors in $\mathbb{C} \times \mathbb{C} \times \mathbb{C} \times T^2$, and is derived from the identification

$$\sum_I \sum_k \lambda^I_{(a)} C_k \phi^I_k \sim \sum_I \sum_k \lambda^I_{(a)} C_k \phi^I_{\sigma(k)} = \sum_I \sum_k \lambda^I_{(a)} C_{\sigma^{-1}(k)} \phi^I_k = e^{iv} \sum_I \sum_k \lambda^I_{(a)} C_k \phi^I_k \,.$$

For $\tau = e^{\pi i/3}$ and $\mathbf{s} = \mathbf{s}''$ we have $v = \frac{\pi}{3}$ and therefore no zero modes for $n < 6$, while for for $\mathbf{s} = -\mathbf{s}''$ we have $v = \frac{4\pi}{3}$ and there are no zero modes if $n < 3$. The analysis of the threshold cases is similar to the one above.

Thus, when we reduce the $\mathcal{N} = 4$ theory to its low-energy limit and then compactify with a twist, we find no low-energy zero-modes for

- $\tau = i$ and $\mathbf{s} = \mathbf{s}'$ if $n = 1, 2, 3$;
- $\tau = e^{\pi i/3}$ and $\mathbf{s} = \mathbf{s}''$ if $n = 1, 2, 3, 4, 5$;
- $\tau = e^{\pi i/3}$ and $\mathbf{s} = -\mathbf{s}''$ if $n = 1, 2$.

We can also provide an alternative argument that does not use the low-energy limit of $\mathcal{N} = 4$ SYM. We can glean more information about the putative Coulomb branch of the low-energy 2+1D theory by studying the VEV of BPS operators. Set $Z \equiv \phi^5 + i\phi^6$ and consider the BPS operators

$$\mathcal{O}_p \equiv g_{\mathrm{YM}}^{-p} \,\mathrm{tr}(Z^p) \,, \qquad p = 1, 2, \ldots$$

According to [9], with this normalization the operators are $\mathrm{SL}(2, \mathbb{Z})$-duality invariant. The action of γ is

$$(\mathcal{O}_p)^\gamma = e^{ipv} \mathcal{O}_p \,.$$

It follows that $\mathcal{O}_p$ is single-valued in our setting if and only if $e^{ipv} = 1$. Therefore,

- for $\tau = i$ and $\mathbf{s} = \mathbf{s}'$, $\langle \mathcal{O}_p \rangle \neq 0$ requires $p \in 4\mathbb{Z}$;
- for $\tau = e^{\pi i/3}$ and $\mathbf{s} = \mathbf{s}''$ $\langle \mathcal{O}_p \rangle \neq 0$ requires $p \in 6\mathbb{Z}$;
- for $\tau = e^{\pi i/3}$ and $\mathbf{s} = -\mathbf{s}''$ $\langle \mathcal{O}_p \rangle \neq 0$ requires $p \in 3\mathbb{Z}$.

For $U(n)$, $\mathcal{O}_{n+1}, \mathcal{O}_{n+2}, \dots$ are not independent of $\mathcal{O}_1, \dots, \mathcal{O}_n$. Thus for $\tau = i$ and $\mathbf{s} = \mathbf{s}'$, for example, if $n < 4$ none of the operators $\mathcal{O}_p$ can get a VEV. By studying additional BPS operators we can restrict the form of the Coulomb branch and reach the same conclusion as above that for $n < 4$ there are no zero modes.

In the rest of this paper, we take the gauge group to have low enough rank so that after S-duality and R-symmetry twists there are no zero modes in the compactified theory.

3. S-duality kernel and topological field theory

In the limit $R \to 0$, the construction in §2 essentially reduces to a question about the duality transformation defined by the $\mathrm{SL}(2,\mathbb{Z})$ element $\mathbf{s}$. In order to see what this question is more precisely, it is convenient to think about the S^1 as a (Euclidean) time direction, and think about the $\mathbf{s}$-twist as an insertion of an operator that realizes the duality. To explore this point of view further, we now switch to a formal Schrödinger representation.

3.1 Schrödinger representation

We will assume that the theory is formulated on a compact three-manifold M_3, so that the full spacetime is $\mathbb{R} \times M_3$. We will use the convention that $i, j, \dots = 1, 2, 3$ are spatial indices and $\mu, \nu, \dots = 0, 1, 2, 3$ are spacetime indices. We will denote the 3D metric on M_3 by g_{ij}. The full metric will be $ds^2 = -dt^2 + g_{ij} dx^i dx^j$. In this subsection, we will not restrict g_{YM} and θ, and work with a generic coupling constant

$$\tau = \frac{4\pi i}{g_{\mathrm{YM}}^2} + \frac{\theta}{2\pi} .$$

We will work in the Hamiltonian formalism and in the temporal gauge

$$A_0 = 0.$$

The spatial components of the gauge field will be denoted by the 1-form $A \equiv A_i dx^i$ that is to be understood as defined on M_3 at a fixed time. Thus, by dA we will always mean "the exterior derivative on M_3" so that dA is a 2-form on M_3. In addition to

the gauge field, we have 6 adjoint-valued scalar fields Φ^I ($I = 1, \ldots, 6$) and 4 spinor fields ψ^a_α ($a = 1, \ldots, 4$ and $\alpha = 1, 2$) and their complex conjugates $\psi_{a\dot{\alpha}}$ ($a = 1, \ldots, 4$ and $\dot{\alpha} = \dot{1}, \dot{2}$). We will denote the collective configuration field by $V \equiv \{A, \psi^a_\alpha, \Phi^I\}$. Physical states are then formally represented by gauge invariant wavefunctions $\Psi(V) \equiv \Psi\{A, \psi^a_\alpha, \Phi^I\}$.

We denote the ($\mathfrak{u}(n)$ or $\mathfrak{su}(n)$ Lie algebra valued) vector field canonically dual to A by $E^i \partial_i$. We understand it as the operator

$$E^i \equiv -2\pi i \frac{\delta}{\delta A_i}$$

acting on wavefunctions. The magnetic field is $B_i dx^i = *dA$, where $*$ is the three-dimensional Hodge star operator. The Hamiltonian is

$$H = \int \sqrt{g} d^3x \, \mathrm{tr}\{\frac{1}{2} g^2_{\mathrm{YM}} g_{ij} E^i E^j + \frac{1}{2g^2_{\mathrm{YM}}} g^{ij} B_i B_j + \cdots\} \ . \tag{3.1}$$

The conjectured S-dual description involves dual fields: $\widetilde{A}_i$, $\widetilde{\psi}^a_\alpha$, $\widetilde{\Phi}^I$, which will be collectively denoted by $\widetilde{V}$. The dual coupling constant and θ-angle are given by

$$\frac{4\pi i}{\tilde{g}^2_{\mathrm{YM}}} + \frac{\widetilde{\theta}}{2\pi} \equiv \widetilde{\tau} = \frac{\mathbf{a}\tau + \mathbf{b}}{\mathbf{c}\tau + \mathbf{d}} \ .$$

The dual Hamiltonian will be denoted by $\widetilde{H}$.

3.2 S-duality kernel

Formally, the Hamiltonian (3.1) acts on the Hilbert space of gauge-invariant wavefunctions. We are interested in how the $\mathrm{SL}(2, \mathbb{Z})$ group of dualities are realized in the Hamiltonian formalism.

The group $\mathrm{SL}(2, \mathbb{Z})$ is generated by $\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ and $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$. The latter corresponds to a shift $\theta \to \theta + 2\pi$ and acts on the wavefunction in a simple way:

$$\Psi(V) \to e^{iI_{CS}(A)} \Psi(V) \, , \tag{3.2}$$

where

$$I_{CS}(A) \equiv \frac{1}{4\pi} \int \mathrm{tr}\{A \wedge dA + \frac{2}{3} A \wedge A \wedge A\} d^3x$$

is the level-1 Chern–Simons action. Equations (3.2) can be seen either by directly integrating the extra $F \wedge F$ term in the action that results from the shift $\theta \to \theta + 2\pi$, or by checking that it acts on the electric field operator in the appropriate way: $E_i \to E_i + B_i$.

S-duality acts on the wavefunction in a more complicated and generally unknown way. We will denote the S-duality operator acting on the Hilbert space by $\widehat{\mathcal{S}}$. We have

$$\widehat{\mathcal{S}}H = \widetilde{H}\widehat{\mathcal{S}}\,. \tag{3.3}$$

Formally, acting on $\Psi(V)$, it produces a dual wavefunction $\widetilde{\Psi}(\widetilde{V}) \equiv \widetilde{\Psi}\{\widetilde{A}, \widetilde{\psi}^a_\alpha, \widetilde{\Phi}^I\}$. We can represent S-duality by a (Fredholm) kernel $\mathcal{S}(V, \widetilde{V})$ that acts as

$$\widetilde{\Psi}(\widetilde{V}) = \int \mathcal{S}(V, \widetilde{V})\Psi(V)[\mathcal{D}V],$$

where $[\mathcal{D}V]$ denotes a path integral on all the fields A, ψ^a_α, Φ^I, and $\mathcal{S}$ is a (generally nonlocal) functional of both field configurations, V and $\widetilde{V}$, which is separately invariant under a gauge transformation of V and a gauge transformation of $\widetilde{V}$.[2]

Let us now discuss the possible ambiguities in our definition of $\mathcal{S}$. The S-duality operator $\widehat{\mathcal{S}}$ transforms eigenstates of the Hamiltonian H to eigenstates of $\widetilde{H}$, but we can consider multiplication by unitary operators that commute with the Hamiltonians. In other words, let Ω be some unitary operator that commutes with H and let $\widetilde{\Omega}$ be another unitary operator that commutes with $\widetilde{H}$. We are considering the freedom to change $\widehat{\mathcal{S}} \mapsto \widetilde{\Omega}\widehat{\mathcal{S}}\Omega$. In the following discussion we will argue that there are essentially no nontrivial ambiguities except for a global phase and time translation. We will use the fact that $\widehat{\mathcal{S}}$ commutes with the Hamiltonian (in the sense of (3.3)) and with R-charge, and takes local operators of $\mathcal{N} = 4$ SYM to local operators of the dual theory.

Since the unitary operator Ω commutes with the Hamiltonian H, it must be a function of the conserved charges which are H and the R-charge operators, and similarly $\widetilde{\Omega}$ must be a function of $\widetilde{H}$ and R-charge of dual theory. Since $\widehat{\mathcal{S}}$ commutes with the 15 R-charge generators, Ω and $\widetilde{\Omega}$ can only depend on $SU(4)$-invariant combinations of the $SU(4)$-generators, i.e., on the Casimirs of $SU(4)$. For a generic positively curved metric g_{ij} and generic $SU(4)_R$ bundle over a compact M_3, we expect the energy levels to fall into $SU(4)_R$ multiplets, but other than that to be discrete and nondegenerate. Thus, for a fixed metric and R-bundle, we may assume that Ω and $\widetilde{\Omega}$ only depend on H and $\widetilde{H}$, respectively, but not on the R-charge. (The Casimirs of the $SU(4)$ representation can be absorbed in a function of the energy, as the representation can generically be read off from the energy.) We can then set $\Omega = f(H)$ and $\widetilde{\Omega} = g(\widetilde{H})$ for some functions f, g of the discrete energy eigenvalues.

[2]The wavefunction Ψ is closely related, after Wick rotation, to the boundary conditions defined in [12]. For example, Dirichlet boundary conditions on the gauge fields correspond to a wavefunction $\delta(B_i)$, and Neumann boundary conditions correspond to a constant wavefunction (on which $\delta/\delta A_i \to 0$). More complicated wavefunctions were constructed in [12] by coupling the gauge degrees of freedom to extra boundary (3D) degrees of freedom. Our $\mathcal{S}$ is directly related to the action of S-duality on these boundary conditions as described in [13].

We now have $\widetilde{\Omega}\widehat{\mathcal{S}}\Omega = \widehat{\mathcal{S}}g(H)f(H)$, and we can assume without loss of generality that $\widetilde{\Omega} = 1$. We also know that $\widehat{\mathcal{S}}$ takes local operators $\mathcal{O}$ to local operators $\widehat{\mathcal{S}}\mathcal{O}\widehat{\mathcal{S}}^{-1}$. To preserve this propetry, $\widehat{\mathcal{S}}\Omega\mathcal{O}\Omega^{-1}\widehat{\mathcal{S}}^{-1}$ must also be local. Here, we are considering locality in time as well as in space. This restricts Ω to be of the form $e^{iTH+i\phi}$ for some constant T and constant phase ϕ. This reflects the obvious fact that there is no canonical way to identify the original time and the time in the dual theory, and so the S-duality has an undetermined time-translation in it.

We can do better when the coupling constant τ takes one of the S-dual values discussed in §2.1 and $\hat{\mathcal{S}}$ realizes the action of corresponding element $\mathbf{s} \in \mathrm{SL}(2,\mathbb{Z})$. Then, we can identify the original Hilbert space with the dual Hilbert space, and also require that the ground state transforms into itself, without a phase. This completely eliminates all the ambiguity.

3.3 Metric independence

We will now argue that $\mathcal{S}$ is topological. To show that, we need to check that $\mathcal{S}$ is independent of the metric g_{ij}. Consider a small deformation δg_{ij} of the metric on M_3. The corresponding corrections to H and $\widetilde{H}$ are

$$\delta H = \int_{M_3} \sqrt{g}\delta g_{ij} T^{ij} d^3x \,, \qquad \delta\widetilde{H} = \int_{M_3} \sqrt{g}\delta g_{ij} \widetilde{T}^{ij} d^3x \,,$$

where T^{ij} are the spatial components of the energy-momentum tensor of the original $\mathcal{N} = 4$ theory, and $\widetilde{T}^{ij}$ are the components of the energy-momentum tensor of the dual theory. Since S-duality maps the energy-momentum tensor to itself,

$$\widehat{\mathcal{S}}(\delta H) = (\delta\widetilde{H})\widehat{\mathcal{S}} \,.$$

Let $\delta\widehat{\mathcal{S}}$ be the small change in $\widehat{\mathcal{S}}$ due to the metric deformation above. Then,

$$(\widehat{\mathcal{S}} + \delta\widehat{\mathcal{S}})(H + \delta H) = (\widetilde{H} + \delta\widetilde{H})(\widehat{\mathcal{S}} + \delta\widehat{\mathcal{S}}) \Longrightarrow (\delta\widehat{\mathcal{S}})H - \widetilde{H}(\delta\widehat{\mathcal{S}}) = 0 \,.$$

Thus, $\delta\widehat{\mathcal{S}}$ transforms an eigenstate of H to an eigenstate of $\widetilde{H}$ with the same energy eigenvalue. Then $\widehat{\mathcal{S}}^{-1}(\widehat{\mathcal{S}}+\delta\widehat{\mathcal{S}})$ commutes with the Hamiltonian. We can similarly argue that it commutes with the R-charge. In addition, if $\mathcal{O}$ is a local operator in the original Hilbert space, then so is $\widehat{\mathcal{S}}^{-1}(\widehat{\mathcal{S}}+\delta\widehat{\mathcal{S}})\mathcal{O}(\widehat{\mathcal{S}}+\delta\widehat{\mathcal{S}})^{-1}\widehat{\mathcal{S}}$. By the same argument as above it then follows that $\widehat{\mathcal{S}}^{-1}(\widehat{\mathcal{S}}+\delta\widehat{\mathcal{S}})$ is the identity operator (up to a possible time translation and a global phase), which proves the claim. A similar argument shows that $\mathcal{S}$ is also independent of the coupling constant τ. To show that, we can use the fact that the dilaton operator $\partial H/\partial\tau$ maps to itself under S-duality.

The construction in §2 is related to the S-kernel $\mathcal{S}$ in the following way. Take one of the selfdual values of τ and the corresponding element $\mathbf{s} \in \mathrm{SL}(2, \mathbb{Z})$ as in §2.1, and let the S-duality and R-symmetry twists be at $x_3 = 0$. Suppose we insert some Wilson-line operator $\mathcal{O}$ at $x_3 = \epsilon > 0$ with $\epsilon \ll R$. Its expectation value in the theory of §2 is given by $\mathrm{tr}\{(-1)^F e^{-2\pi RH} \mathcal{O}\widehat{\mathcal{S}}\gamma\}$, where $\hat{\mathcal{S}}$ represents the action of $\mathbf{s}$ on the Hilbert space. If we could take the naive limit $R \to 0$, we would get $\mathrm{tr}\{(-1)^F \mathcal{O}\widehat{\mathcal{S}}\gamma\}$ which corresponds to calculating the expectation value of $\mathcal{O}$ in a 3D theory whose action is formally given by

$$\mathcal{I}(V) \equiv -i \log \mathcal{S}(V, V^\gamma)\,, \tag{3.4}$$

where V^γ are the fields V after the R-symmetry twist γ. In other words, the expectation value of an operator $F(V)$ (say a Wilson line) in that theory is by definition given by

$$\langle F \rangle \equiv \int e^{i\mathcal{I}(V)} F(V)[\mathcal{D}V] \equiv \int \mathcal{S}(V, V^\gamma) F(V)[\mathcal{D}V]\,.$$

However, since H is not bounded from above, taking the limit $R \to 0$ is potentially dangerous. One potential problem is that $\mathrm{tr}\{(-1)^F \mathcal{O}\widehat{\mathcal{S}}\gamma\}$ receives large contributions from high-energy modes, for example if $\mathcal{O}$ is a Wilson line with a cusp. Another problem is if $\mathrm{tr}\{(-1)^F \mathcal{O}\widehat{\mathcal{S}}\gamma\}$ is ill-defined because of massless zero-modes. This is the case if γ is not generic enough. For the time being, we will assume that none of these problems arise and that the $R \to 0$ limit is safe. We will proceed to discuss the diagonal of the S-duality kernel, as defined in (3.4).

In general, the kernel $\mathcal{S}$ depends on two independent field configurations V and $\widetilde{V}$ on M_3, and we do not expect it to be expressible in terms of an integral of a local expression in the fields. However, the construction in §2 makes it clear that $\mathcal{I}(V)$ is the action of a local theory. We further conjecture that $\mathcal{I}(V)$ is an integral of a local expression in the fields V, i.e., the theory that it defines is not only local, but local in the variables V. Since we argued that $\mathcal{S}(V, \widetilde{V})$ is topological, i.e., independent of the metric g_{ij}, we expect $\mathcal{I}(V)$ to also be topological. We conclude that the diagonal of the S-duality kernel defines a local topological field theory in three dimensions.

3.4 Expectation value of a Wilson-loop pair

An interesting aspect of the nonlocal structure $\mathcal{S}(V, \widetilde{V})$ is that it renders the correlation functions of Wilson loops easy to calculate. For this purpose, we specialize to the $\tau = i$ case, and let the nonlocal kernel $\mathcal{S}(V, \widetilde{V})$ represent the action of $\mathbf{s} = \mathbf{s}' \in \mathrm{SL}(2, \mathbb{Z})$. For clarity of discussion, we momentarily suppress the dependence on superpartners and write the kernel as $\mathcal{S}(A, \widetilde{A})$.

Choose two loops $\mathbf{C}$ and $\widetilde{\mathbf{C}}$ in a three-manifold and define the double-Wilson-loop expectation value:

$$W(\mathbf{C}, \widetilde{\mathbf{C}}) = \int \mathcal{S}(A, \widetilde{A}) \operatorname{tr}\left(P e^{i \oint_{\mathbf{C}} A}\right) \operatorname{tr}\left(P e^{i \oint_{\widetilde{\mathbf{C}}} \widetilde{A}}\right) [\mathcal{D}A][\mathcal{D}\widetilde{A}] \,,$$

where tr is in the fundamental representation of $SU(n)$. We will now present a heuristic calculation of $W(\mathbf{C}, \widetilde{\mathbf{C}})$. We work in the formal Schrödinger representation discussed in §3.1 and define the operators

$$\widehat{W}(\mathbf{C}) = \operatorname{tr}\left(P e^{i \oint_{\mathbf{C}} A}\right) \,, \qquad \widehat{W}(\widetilde{\mathbf{C}}) = \operatorname{tr}\left(P e^{i \oint_{\widetilde{\mathbf{C}}} A}\right) \,.$$

We also define the formal state $|1\rangle$ which has a formal wavefunctional $\Psi\{A\} = 1$ for every gauge field configuration A (which is related to Neumann boundary conditions in [12]). Then, formally,

$$W(\mathbf{C}, \widetilde{\mathbf{C}}) = \langle 1|\widehat{W}(\widetilde{\mathbf{C}})\mathcal{S}\widehat{W}(\mathbf{C})|1\rangle \,.$$

Let $\widehat{M}(\widetilde{\mathbf{C}})$ be the 't Hooft loop operator [33] associated with the loop $\widetilde{\mathbf{C}}$. Then $\widehat{W}(\widetilde{\mathbf{C}})\mathcal{S} = \mathcal{S}\widehat{M}(\widetilde{\mathbf{C}})$ (see [10][21]). Using the commutation relation [33]

$$\widehat{M}(\widetilde{\mathbf{C}})\widehat{W}(\mathbf{C}) = \widehat{W}(\mathbf{C})\widehat{M}(\widetilde{\mathbf{C}}) e^{\frac{2\pi i}{n} L(\mathbf{C}, \widetilde{\mathbf{C}})} \,,$$

where $L(\mathbf{C}, \widetilde{\mathbf{C}})$ is the *linking number* of the loops $\mathbf{C}$ and $\widetilde{\mathbf{C}}$, and using the fact that $\widehat{M}(\widetilde{\mathbf{C}})$ acts by changing one gauge configuration to another, so that $\widehat{M}(\widetilde{\mathbf{C}})|1\rangle = |1\rangle$, we get

$$\begin{aligned} W(\mathbf{C}, \widetilde{\mathbf{C}}) &= \langle 1|\widehat{W}(\widetilde{\mathbf{C}})\mathcal{S}\widehat{W}(\mathbf{C})|1\rangle = \langle 1|\mathcal{S}\widehat{M}(\widetilde{\mathbf{C}})\widehat{W}(\mathbf{C})|1\rangle = e^{\frac{2\pi i}{n} L(\mathbf{C}, \widetilde{\mathbf{C}})} \langle 1|\mathcal{S}\widehat{W}(\mathbf{C})\widehat{M}(\widetilde{\mathbf{C}})|1\rangle \\ &= e^{\frac{2\pi i}{n} L(\mathbf{C}, \widetilde{\mathbf{C}})} \langle 1|\mathcal{S}\widehat{W}(\mathbf{C})|1\rangle = e^{\frac{2\pi i}{n} L(\mathbf{C}, \widetilde{\mathbf{C}})} \langle 1|\widehat{M}(\mathbf{C})\mathcal{S}|1\rangle = e^{\frac{2\pi i}{n} L(\mathbf{C}, \widetilde{\mathbf{C}})} \langle 1|\mathcal{S}|1\rangle \,, \end{aligned}$$

The last equality is justified by inserting $\int [\mathcal{D}A] |A\rangle\langle A|$ in front of $\mathcal{S}$ on both sides and using $\langle 1|\widehat{M}(\mathbf{C})|A\rangle = 1 = \langle 1|A\rangle$ for any configuration eigenstate $|A\rangle$, according to the definition of $|1\rangle$.

The normalization factor $\langle 1|\mathcal{S}|1\rangle$ is independent of the loops $\mathbf{C}, \widetilde{\mathbf{C}}$, and assuming that it can be regularized to a nonzero value, we find a topological result:

$$W(\mathbf{C}, \widetilde{\mathbf{C}}) \propto e^{\frac{2\pi i}{n} L(\mathbf{C}, \widetilde{\mathbf{C}})} \,. \tag{3.5}$$

In principle, this result can be used to reconstruct the nonlocal kernel $\mathcal{S}(A, \widetilde{A})$, at least on a lattice. It would be interesting to see if this can lead to a useful expression for $\mathcal{S}$ [34].

3.5 Electric and magnetic fluxes

We close this section by studying how the S-duality operator $\hat{\mathcal{S}}$ acts on the electric and magnetic fluxes [33]. Our analysis is based on a review contained in [21], to which we refer the reader for more information.

Let us start with the gauge group $SU(n)$ and its adjoint form $SU(n)/\mathbb{Z}_n$. Since we consider the theory on a manifold $X = S^1 \times M_3$, where we view the S^1 as Euclidean time, the electric and magnetic fluxes $\mathbf{e}$ and $\mathbf{m}$ take values in the following abelian groups:

$$\mathbf{e} \in \mathrm{Hom}(H^1(M_3, \mathbb{Z}_n), U(1)), \qquad \mathbf{m} \in H^2(M_3, \mathbb{Z}_n)\,. \tag{3.6}$$

As $\mathrm{Hom}(H^1(M_3; \mathbb{Z}_n), U(1))$ is naturally isomorphic to $H^2(M_3; \mathbb{Z}_n)$, one can meaningfully talk about exchanging the electric and magnetic fluxes. More precisely, the S-duality conjecture states that the fluxes transform as

$$\begin{pmatrix} \mathbf{e} \\ \mathbf{m} \end{pmatrix} \to \begin{pmatrix} \tilde{\mathbf{e}} \\ \tilde{\mathbf{m}} \end{pmatrix} = \begin{pmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{pmatrix} \begin{pmatrix} \mathbf{e} \\ \mathbf{m} \end{pmatrix} \tag{3.7}$$

under the $\mathrm{SL}(2, \mathbb{Z})$ action.

Each choice of $\mathbf{e}$ and $\mathbf{m}$ defines a Hilbert space $\mathcal{H}_{\mathbf{e},\mathbf{m}}$. Its elements are those wavefunctions defined on the space of $SU(n)/\mathbb{Z}_n$ bundles of topological type defined by $\mathbf{m}$ that transform in a way specified by $\mathbf{e}$ under a large gauge transformation. In view of (3.7), the S-duality operator $\hat{\mathcal{S}}$ acts on these Hilbert spaces as

$$\hat{\mathcal{S}} : \mathcal{H}_{\mathbf{e},\mathbf{m}} \to \mathcal{H}_{\tilde{\mathbf{e}},\tilde{\mathbf{m}}}\,. \tag{3.8}$$

Therefore, with the choices of gauge coupling τ and $\mathbf{s} \in \mathrm{SL}(2, \mathbb{Z})$ that we have considered so far, $\mathcal{H}_{\mathbf{e},\mathbf{m}}$ is invariant under duality in the following cases[3]:

- for $\tau = i$ and $\mathbf{s} = \mathbf{s}'$: $\mathbf{e} = \mathbf{m}$ and $2\mathbf{m} = 0$;
- for $\tau = e^{\pi i/3}$ and $\mathbf{s} = \mathbf{s}''$: $\mathbf{e} = \mathbf{m} = 0$;
- for $\tau = e^{\pi i/3}$ and $\mathbf{s} = -\mathbf{s}''$: $\mathbf{e} + \mathbf{m} = 0$ and $3\mathbf{e} = 0$.

For example, if $n = 2$, then $2\mathbf{m} = 0$ for any $\mathbf{m}$, so at $\tau = i$, the Hilbert space $\mathcal{H}_{\mathbf{e},\mathbf{e}}$ is invariant for any $\mathbf{e}$ under the action of $\mathbf{s} = \mathbf{s}'$.

In the $U(1)$ theory, the electric and magnetic fluxes take values in $H^2(M_3, \mathbb{Z})$, and also transform as (3.7) under the duality. It is easy to see that the only invariant Hilbert space $\mathcal{H}_{\mathbf{e},\mathbf{m}}$ in this case is the one with $\mathbf{e} = \mathbf{m} = 0$.

[3]Note that $\mathbf{e}, \mathbf{m}$ are elements of abelian torsion groups where $n \cdot \mathbf{m} = 0$ and $n \cdot \mathbf{e} = 0$ doesn't imply $\mathbf{m} = 0$ and $\mathbf{e} = 0$.

By combining the results for the $SU(n)$ and $U(1)$ cases above, we can answer the same question for the $U(n)$ theory. We first note that the Hilbert space for the $U(n)$ theory decomposes as

$$\mathcal{H}^{U(n)} = \bigoplus_{\substack{\mathbf{e}'=\mathbf{e} \mod n \\ \mathbf{m}'+\mathbf{m}=0 \mod n}} \mathcal{H}^{U(1)}_{\mathbf{e}',\mathbf{m}'} \otimes \mathcal{H}^{SU(n)}_{\mathbf{e},\mathbf{m}} \,. \tag{3.9}$$

The fluxes are correlated because of the $\mathbb{Z}_n$ action in $U(n) = [SU(n) \times U(1)]/\mathbb{Z}_n$. It follows from the above results that, for all values of τ and $\mathbf{s}$ under consideration, the only invariant Hilbert space $\mathcal{H}^{U(1)}_{\mathbf{e}',\mathbf{m}'} \otimes \mathcal{H}^{SU(n)}_{\mathbf{e},\mathbf{m}}$ is the one with $(\mathbf{e}', \mathbf{m}') = (0, 0)$ and $(\mathbf{e}, \mathbf{m}) = (0, 0)$.

Now, let us return to $SU(n)$. We have seen above that only a small subset of the possible flux combinations $\mathbf{e}, \mathbf{m}$ are allowed. How can we modify the $\mathrm{SL}(2,\mathbb{Z})$-twist construction to include fluxes that are not $\mathrm{SL}(2,\mathbb{Z})$ invariant? Suppose the $\mathrm{SL}(2,\mathbb{Z})$-twist is at $x_3 = 0$ and that for $x_3 < 0$ we have fluxes $\mathbf{e}, \mathbf{m}$, so that for $x_3 > 0$ we have fluxes $\tilde{\mathbf{e}}, \tilde{\mathbf{m}}$, as in (3.7). To make this construction consistent we need to insert an operator at some other x_3, say $x_3 = \epsilon$, that augments the electric flux by $\mathbf{e} - \tilde{\mathbf{e}}$ and augments the magnetic flux by $\mathbf{m} - \tilde{\mathbf{m}}$. Let $\mathbf{C}$ be a loop (or union of loops) in M_3 whose homology class is equivalent to the cohomology class $\mathbf{e} - \tilde{\mathbf{e}}$. Then, an appropriate Wilson loop operator $W(\mathbf{C})$ inserted at $x_3 = \epsilon$ is the operator we need. Similarly, a 't Hooft loop for $\mathbf{C}$ that is Poincaré dual to the cohomology class $\mathbf{m} - \tilde{\mathbf{m}}$ will augment the magnetic flux by the desired amount.

For our application we especially need to consider the case $M_3 = \mathcal{C}_h \times \mathbb{R}$, where $\mathcal{C}_h$ is some Riemann surface (of genus h). Let $\mathbf{e}_0, \mathbf{m}_0$ be the electric and magnetic fluxes through $\mathcal{C}_h$, which take values in $\mathbb{Z}_n$. Choose a point $p \in \mathcal{C}_h$ and a representation $\mathbf{r}$ of $SU(n)$ and consider the Wilson-loop operator

$$W(\mathbf{r}, p, \epsilon) = \mathrm{tr}_{\mathbf{r}} \left(P \exp \oint_{\{p\}\times\mathbb{R}\times\{x_3=\epsilon\}} A \right).$$

Here, $\{p\} \times \mathbb{R} \times \{x_3 = \epsilon\} \subset \mathcal{C}_h \times \mathbb{R} \times S^1$ is the line at p and $x_3 = \epsilon$. This operator $W(\mathbf{r}, p, \epsilon)$ augments the electric flux by the number of boxes b of the Young diagram associated with $\mathbf{r}$. Thus, we require $b \equiv (\mathbf{e}_0 - \tilde{\mathbf{e}}_0) \mod n$, where $\tilde{\mathbf{e}}_0$ is the electric flux through $\mathcal{C}_h$ after the $\mathrm{SL}(2,\mathbb{Z})$-duality $\mathbf{s}$. Similarly, a 't Hooft loop at p will suffice if $b \equiv (\mathbf{m}_0 - \tilde{\mathbf{m}}_0) \mod n$.

We do not know if there are any further restrictions required of $\mathbf{r}$ in either the electric or magnetic case. We therefore believe, for example, that the compactification $\mathcal{C}_h \times \mathbb{R} \times S^1$ with the $\mathbf{s}'$ twist at $x_3 = 0$ in the S^1 direction, with $\mathbf{m}_0 = \mathbf{e}_0 = m$ for $0 < x_3 < \epsilon$ (and some $m \in \mathbb{Z}_n$) and with $\mathbf{m}_0 = -\mathbf{e}_0 = m$ for $\epsilon < x_3 < 2\pi$, and with

a straight Wilson line at $x_3 = \epsilon$ and $p \in \mathcal{C}_h$ in a representation $\mathbf{r}$ with $2m$ boxes, is consistent.

4. T-duality twist and geometric quantization

Before we approach the main case of interest, $\mathcal{N} = 4$ SYM with an S-duality twist, it is useful to study a simpler problem where similar ideas arise. The problem that we will study in this section is the compactification of a 1+1D theory on S^1 with the insertion of a T-duality twist, assuming the theory is selfdual. Arguments along the line presented in §3 suggest that this construction yields a topological theory in 0+1D. A 0+1D topological theory is a quantum mechanical system with a Hamiltonian that is identically zero. In the examples that we study below, it will have a finite-dimensional Hilbert space (of vacua).

We will begin with a free scalar at the self-dual radius. In this case, we will see that the resulting topological theory is trivial. We will then proceed to a free σ-model with target space T^d at a point in the moduli space that is invariant under some duality transformation in $O(d, d, \mathbb{Z})$. We will demonstrate that the resulting topological theory is equivalent to geometric quantization of the target space. We will then comment on a more general case where we twist a selfdual supersymmetric σ-model by the duality that is mirror symmetry.

4.1 Warm-up: free self-dual scalar

Consider a free real 1+1D boson $\Phi(\tilde{\sigma}, \tilde{\tau})$ where $0 \le \tilde{\sigma} < 2\pi$ is the spatial coordinate and $\tilde{\tau}$ is time. The action is

$$S = \frac{1}{4\pi} \int \left\{ (\partial_{\tilde{\tau}} \Phi)^2 - (\partial_{\tilde{\sigma}} \Phi)^2 \right\} d\tilde{\sigma} d\tilde{\tau} \,,$$

and the Hamiltonian is

$$H = \frac{1}{4\pi} \int \left\{ (\partial_{\tilde{\tau}} \Phi)^2 + (\partial_{\tilde{\sigma}} \Phi)^2 \right\} d\tilde{\sigma} \,.$$

We take the boson at the selfdual radius, so that $\Phi \sim \Phi + 2\pi$.

There are many ways to prove T-duality of this simple free theory [35], but for our purposes we need to do it in the Schrödinger representation. We therefore expand, at fixed τ,

$$\Phi(\tilde{\sigma}) = w\tilde{\sigma} + \sum_{n=-\infty}^{\infty} \phi_n e^{in\tilde{\sigma}} \,,$$

where $w \in \mathbb{Z}$ is the winding number, $\phi_n^* = \phi_{-n}$ are the Fourier modes, and ϕ_0 is real and periodic with period 2π. A state in the Hilbert space is described by a wavefunction, which is a formal expression $\Psi(w, \{\phi_n\})$, and T-duality acts as

$$\Psi(w, \{\phi_n\}) \to \widetilde{\Psi}(\tilde{w}, \{\tilde{\phi}_n\}) = \sum_w \int \prod_n d\phi_n \mathcal{T}(\tilde{w}, \{\tilde{\phi}_n\}; w, \{\phi_n\}) \Psi(w, \{\phi_n\}) \,,$$

where the duality kernel $\mathcal{T}$ is given by [38]:

$$\begin{aligned} \mathcal{T}(\tilde{w}, \{\tilde{\phi}_n\}; w, \{\phi_n\}) &= \exp\{i(\tilde{w}\phi_0 - w\tilde{\phi}_0) + \sum_{n=-\infty}^{\infty} n\phi_n\tilde{\phi}_{-n}\} \\ &= \exp\Big\{-iw\tilde{\Phi}(0) - \pi i\tilde{w}w + \frac{i}{2\pi}\int_0^{2\pi} \Phi(\tilde{\sigma})\tilde{\Phi}'(\tilde{\sigma})d\tilde{\sigma}\Big\} \,. \end{aligned} \tag{4.1}$$

(The first term on the second line is required to make the entire expression independent of the choice of origin on the $\tilde{\sigma}$ direction.) This can be checked by noting that this map acts on operators as

$$\partial_{\tilde{\sigma}}\Phi(\tilde{\sigma}_0) \to -2\pi i \frac{\delta}{\delta\Phi(\tilde{\sigma}_0)} \,, \qquad -2\pi i \frac{\delta}{\delta\Phi(\tilde{\sigma}_0)} \to -\partial_{\tilde{\sigma}}\Phi(\tilde{\sigma}_0) \,.$$

Note, however, that we have the freedom to multiply the operator by an arbitrary function of the conserved charges, which are the winding number w and the momentum $p \equiv -i\partial/\partial\phi_0$. The latter has the following interpretation. Consider first an exponential function e^{ipa}, where a is some constant. This function acts by shifting $\phi_0 \to \phi_0 + a$, and the effect on $\mathcal{T}$ in (4.1) is to replace every ϕ_0 with $(\phi_0 + a)$. Now, if we average this over various a's with some weight function $f(a)$, the effect on $\mathcal{T}$ would be to multiply it by some function of $\tilde{w}$. Therefore, the ambiguity in $\mathcal{T}$ can be rephrased as the freedom to multiply by an arbitrary phase that depends on $(w, \tilde{w})$ alone. We will ignore this ambiguity and take (4.1) as the expression that defines the duality kernel.

Based on what we learned in §3.3, we expect that the T-duality kernel $\mathcal{T}$ defines a topological theory in one less dimension when we equate the original variables w and ϕ_n to their dual partners $\tilde{w}$ and $\tilde{\phi}_n$. Setting $\tilde{w} = w$ and $\tilde{\phi}_n = \phi_n$ in (4.1), we find that the diagonal of $\mathcal{T}$ is identically zero—this is a trivial topological theory. Note also that the second line of (4.1) is a topological expression (independent of the 0+1D metric). Interpreting $\tilde{\sigma}$ as time, the discussion above implies that the Hilbert space of our topological theory has only one state. We now switch the role of $\tilde{\sigma}$ and $\tilde{\tau}$, and from now on, unless otherwise stated, we interpret $\tilde{\sigma}$ as spatial and $\tilde{\tau}$ as temporal.

4.2 T^d target space

Our first nontrivial (yet simple) example is a σ-model with target space T^2 that is a product of two circles $S^1 \times S^1$, one with radius R_1 and the other with radius $R_2 = 1/R_1$. The action is

$$S = \sum_{k=1}^{2} \frac{1}{4\pi R_k^2} \int \left\{ (\partial_{\tilde{\tau}} \Phi_k)^2 - (\partial_{\tilde{\sigma}} \Phi_k)^2 \right\} d\tilde{\sigma} d\tilde{\tau} \,, \tag{4.2}$$

and the theory is selfdual under a simultaneous T-duality in both directions, combined with an exchange of the two S^1's. In order to get a nontrivial result, it turns out that we need to add to the twist a reflection in one of the S^1's. With this reflection, the combined duality also preserves the complex structure of the T^2.

For the Schrödinger formalism, we expand

$$\Phi_k(\tilde{\sigma}) = w^{(k)} \tilde{\sigma} + \sum_{n=-\infty}^{\infty} \phi_n^{(k)} e^{in\tilde{\sigma}} \,, \qquad k = 1, 2,$$

where $\phi_0^{(k)}$ are real and periodic with period 2π. A state in the Hilbert space is described by a formal wavefunction $\Psi(w^{(1)}, w^{(2)}, \{\phi_n^{(k)}\})$, and T-duality acts as

$$\begin{aligned} \Psi(\{w^{(k)}\}, \{\phi_n^{(k)}\}) &\rightarrow \widetilde{\Psi}(\tilde{w}^{(k)}, \{\tilde{\phi}_n^{(k)}\}) \\ &= \sum_{w^{(1)}, w^{(2)}} \int \prod_n d\phi_n^{(k)} \mathcal{T}(\{\tilde{w}^{(k)}\}, \{\tilde{\phi}_n\}; \{w^{(k)}\}, \{\phi_n\}) \Psi(\{w^{(k)}\}, \{\phi_n^{(k)}\}) \,, \end{aligned}$$

where the duality kernel $\mathcal{T} = \mathcal{T}(\{\tilde{w}^{(k)}\}, \{\tilde{\phi}_n\}; \{w^{(k)}\}, \{\phi_n\})$ is given by

$$\begin{aligned} \mathcal{T} = \exp &\Big\{ i(\tilde{w}^{(2)} \phi_0^{(1)} - w^{(1)} \tilde{\phi}_0^{(2)}) - i(\tilde{w}^{(1)} \phi_0^{(2)} + w^{(2)} \tilde{\phi}_0^{(1)}) \\ &+ \sum_{n=-\infty}^{\infty} n \phi_n^{(1)} \tilde{\phi}_{-n}^{(2)} - \sum_{n=-\infty}^{\infty} n \phi_n^{(2)} \tilde{\phi}_{-n}^{(1)} \Big\} \\ = \exp &\Big\{ -i w^{(1)} \tilde{\Phi}_2(0) - \pi i \tilde{w}^{(2)} w^{(1)} + i w^{(2)} \tilde{\Phi}_1(0) + \pi i \tilde{w}^{(1)} w^{(2)} \\ &+ \frac{i}{2\pi} \int_0^{2\pi} \tilde{\Phi}_2 \partial_{\tilde{\sigma}} \Phi_1 d\tilde{\sigma} - \frac{i}{2\pi} \int_0^{2\pi} \tilde{\Phi}_1 \partial_{\tilde{\sigma}} \Phi_2 d\tilde{\sigma} \Big\} \,. \end{aligned} \tag{4.3}$$

Now we set $w^{(k)} = \tilde{w}^{(k)}$ and $\Phi_k = \tilde{\Phi}_k$ in (4.3), and find that the diagonal of the duality kernel becomes a topological action

$$S = \frac{1}{2\pi} \int_0^{2\pi} \epsilon^{kl} \Phi_k \partial_{\tilde{\sigma}} \Phi_l d\tilde{\sigma} - \frac{1}{2\pi} \epsilon^{kl} \Phi_k(0) \int_0^{2\pi} \partial_{\tilde{\sigma}} \Phi_l d\tilde{\sigma} \,. \tag{4.4}$$

where $\epsilon^{12} = -\epsilon^{21} = 1$ and $\epsilon^{11} = \epsilon^{22} = 0$. We then treat $\tilde{\sigma}$ as a time coordinate, which turns (4.4) into a 0+1D action. The second term on the right-hand side is nonlocal,

but is required in order to make the integral independent of the choice of origin. If desired, we can eliminate this term by replacing the integration range $0 \le \tilde{\sigma} < 2\pi$ with $-\infty < \tilde{\sigma} < \infty$ and taking as a boundary condition $\Phi_k(-\infty) = 0$. The resulting action is then simply

$$S = \frac{1}{2\pi}\int_{-\infty}^{\infty} \epsilon^{kl}\Phi_k\partial_{\tilde{\sigma}}\Phi_l d\tilde{\sigma}\,. \tag{4.5}$$

This action describes geometric quantization of the target space T^2 with a symplectic form

$$\omega = \frac{2}{(2\pi)^2}d\phi_1 \wedge d\phi_2\,,$$

where $0 \le \phi_1 < 2\pi$ and $0 \le \phi_2 < 2\pi$ are coordinates on T^2. With this symplectic form, the area of the target space is 2 and there are therefore two quantum states in the Hilbert space of this simple topological 0+1D theory. In §4.3 we will present a more geometrical description of these two states, and the distinction between them.

Let us now generalize the discussion to a torus T^d of an arbitrary *even* dimension d, with an arbitrary flat metric $G_{IJ}d\phi^I d\phi^J$ and antisymmetric B-field $B_{IJ}d\phi^I \wedge d\phi^J$ ($I, J = 1, \dots, d$). For the duality twist, we pick an element $\mathbf{t}$ in the duality group $O(d, d, \mathbb{Z})$ and write it in block form as

$$\mathbf{t} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix} \in O(d, d, \mathbb{Z})\,,$$

where $\mathbf{A}, \mathbf{B}, \mathbf{C}, \mathbf{D}$ are $d \times d$ matrices. The action on G_{IJ} and B_{IJ} is conveniently expressed as follows (see [36] for a review). Define the $d \times d$ matrix E by

$$E_{IJ} = G_{IJ} + B_{IJ}\,.$$

Then $\mathbf{t}$ acts as

$$E \to (\mathbf{A}E + \mathbf{B})(\mathbf{C}E + \mathbf{D})^{-1}\,.$$

For our purposes, we pick a selfdual background for which

$$E = (\mathbf{A}E + \mathbf{B})(\mathbf{C}E + \mathbf{D})^{-1}\,.$$

The duality acts on the left-moving and right-moving free fields of the σ-model as

$$\partial_-\Phi \to (\mathbf{D} - \mathbf{C}E^t)^{-1}\partial_-\Phi\,, \qquad \partial_+\Phi \to (\mathbf{D} + \mathbf{C}E)^{-1}\partial_+\Phi\,, \qquad \partial_\pm \equiv \partial_{\tilde{\tau}} \pm \partial_{\tilde{\sigma}}\,, \tag{4.6}$$

where Φ here is understood as a d-component vector. (See (2.4.36) of [36].) From this action we calculate

$$\partial_{\tilde{\tau}}\Phi = \frac{1}{2}(\partial_+\Phi + \partial_-\Phi) \to \frac{1}{2}((\mathbf{D} + \mathbf{C}E)^{-1}\partial_+\Phi + (\mathbf{D} - \mathbf{C}E^t)^{-1}\partial_-\Phi) = \mathbf{U}\partial_{\tilde{\tau}}\Phi + \mathbf{V}\partial_{\tilde{\sigma}}\Phi\,,$$
$$\partial_{\tilde{\sigma}}\Phi = \frac{1}{2}(\partial_+\Phi - \partial_-\Phi) \to \frac{1}{2}((\mathbf{D} + \mathbf{C}E)^{-1}\partial_+\Phi - (\mathbf{D} - \mathbf{C}E^t)^{-1}\partial_-\Phi) = \mathbf{U}\partial_{\tilde{\sigma}}\Phi + \mathbf{V}\partial_{\tilde{\tau}}\Phi\,,$$

where

$$\mathbf{U} = (\mathbf{D} + \mathbf{C}E)^{-1}(\mathbf{D} + \mathbf{C}B)(\mathbf{D} - \mathbf{C}E^t)^{-1}\,,$$
$$\mathbf{V} = -(\mathbf{D} + \mathbf{C}E)^{-1}\mathbf{C}G(\mathbf{D} - \mathbf{C}E^t)^{-1}\,.$$

Taking the range $-\infty < \tilde{\sigma} < \infty$ and boundary conditions $\Phi(-\infty) = 0$, to avoid complications, we get (up to total derivatives) the T-duality kernel

$$\mathcal{T} = \exp\left\{\frac{i}{4\pi}\int_{-\infty}^{\infty}\Big(\Phi(2\mathbf{X})\partial_{\tilde{\sigma}}\Phi + \tilde{\Phi}(2\mathbf{Y})\partial_{\tilde{\sigma}}\tilde{\Phi} + \tilde{\Phi}(2\mathbf{Z})\partial_{\tilde{\sigma}}\Phi\Big)d\tilde{\sigma}\right\}, \tag{4.7}$$

where

$$2\mathbf{X} = -B + G\mathbf{V}^{-1}\mathbf{U}\,, \quad 2\mathbf{Y} = B + G\mathbf{U}\mathbf{V}^{-1}\,, \quad \mathbf{Z}^t = G\mathbf{V}^{-1}\,. \tag{4.8}$$

This is found by solving[4]

$$\left(-2\pi i G^{-1}\frac{\delta}{\delta\tilde{\Phi}} - G^{-1}B\partial_{\tilde{\sigma}}\tilde{\Phi}\right)\mathcal{T} = \mathbf{U}\left(2\pi i G^{-1}\frac{\delta}{\delta\Phi} - G^{-1}B\partial_{\tilde{\sigma}}\Phi\right)\mathcal{T} + \mathbf{V}\partial_{\tilde{\sigma}}\Phi\mathcal{T}\,,$$
$$\partial_{\tilde{\sigma}}\tilde{\Phi}\mathcal{T} = \mathbf{U}\partial_{\tilde{\sigma}}\Phi\mathcal{T} + \mathbf{V}\left(2\pi i G^{-1}\frac{\delta}{\delta\Phi} - G^{-1}B\partial_{\tilde{\sigma}}\Phi\right)\mathcal{T}\,.$$

After a little algebra, (4.8) can be simplified as

$$2\mathbf{X} = -E + (\mathbf{C}^t)^{-1}(\mathbf{D} + \mathbf{C}E)^{-1}\,, \quad 2\mathbf{Y} = -\mathbf{C}^{-1}\mathbf{D}\,, \quad \mathbf{Z} = \mathbf{C}^{-1}\,. \tag{4.9}$$

Setting $\Phi = \tilde{\Phi}$ in (4.7), we get the topological action

$$S = \frac{1}{4\pi}\int_{-\infty}^{\infty}\Phi(2\mathbf{W})\partial_{\tilde{\sigma}}\Phi d\tilde{\sigma}\,, \tag{4.10}$$

where

$$2\mathbf{W} = 2\mathbf{X} + 2\mathbf{Y} + \mathbf{Z} - \mathbf{Z}^t = (\mathbf{C}^t)^{-1}(\mathbf{D} + \mathbf{C}E)^{-1} - \mathbf{C}^{-1}(\mathbf{D} + \mathbf{C}E) + \mathbf{C}^{-1} - (\mathbf{C}^t)^{-1}\,.$$

This describes geometric quantization of the target space T^d with the symplectic form given by

$$\omega = \frac{\mathbf{W}_{IJ}}{(2\pi)^d}d\phi^I \wedge d\phi^J\,.$$

For example, we can recover the previous case with T^2 target space by setting

$$G = \begin{pmatrix} R_1^2 & 0 \\ 0 & R_1^{-2} \end{pmatrix}, \quad B = 0\,, \quad \mathbf{A} = \mathbf{D} = 0\,, \quad \mathbf{B} = \mathbf{C} = J \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

[4]In manipulating the matrices here and below, it is essential to note the following properties of $O(d,d,\mathbb{Z})$ matrices: $\mathbf{A}^t\mathbf{C} + \mathbf{C}^t\mathbf{A} = 0$, $\mathbf{B}^t\mathbf{D} + \mathbf{D}^t\mathbf{B} = 0$, $\mathbf{A}^t\mathbf{D} + \mathbf{C}^t\mathbf{B} = 1$.

The matrix J is so chosen as to incorporate the exchange of two circles and reflection in one of them. The number of states in the Hilbert space is

$$n = \mathrm{Pf}(2\mathbf{W}) \,,$$

where Pf denotes the Pfaffian.

Let us specialize again to the case of T^2. In the study of T-duality for T^2, one usually defines the complex combination

$$\rho = B_{12} + i\sqrt{\det G} \,.$$

The duality group $O(2,2,\mathbb{Z})$ is essentially two copies of $\mathrm{SL}(2,\mathbb{Z})$, one acting on ρ and the other acting on the complex structure of T^2 in a geometrical way. Under an element

$$\begin{pmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{pmatrix}$$

in the first $\mathrm{SL}(2,\mathbb{Z})$ factor, ρ transforms as as

$$\rho \to \frac{\mathbf{a}\rho + \mathbf{b}}{\mathbf{c}\rho + \mathbf{d}} \,.$$

We now have three possibilities for a duality twist $\mathbf{t}$, which are analogous to the list in §2.1:

1. $\rho = i$ and $\mathbf{t} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, with 2 vacua;

2. $\rho = e^{\pi i/3}$ and $\mathbf{t} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$, with 1 vacuum;

3. $\rho = e^{\pi i/3}$ and $\mathbf{t} = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$, with 3 vacua.

4.3 An alternative way of counting vacua

We will now describe a more geometrical interpretation for the vacua of the topological 0+1D theories that we obtained in §4.2. We will concentrate on the simple T^2 target space with action (4.2). The trick is simple: perform T-duality only on one of the two circles, say the one corresponding to Φ_2. We now have two circles of equal radius R_1 and a target space T^2 with complex structure $\tau = i$. In this picture selfduality is a geometrical isometry of T^2, which in the realization of T^2 as a lattice $\mathbb{C}/(\mathbb{Z} + \tau\mathbb{Z})$ corresponds to rotation of $\mathbb{C}$ by $\pi/2$. A similar duality can be applied for the other

selfdual values of ρ (which becomes the complex structure τ after T-duality only on one circle) from the list at the end of §4.2. We will proceed with a general τ.

We take the 1+1D coordinates to be $(\tilde{\sigma}, \tilde{\tau})$, and the twist will be at $\tilde{\sigma} = 0 \sim 2\pi$. Set

$$\mathcal{Z}(\tilde{\sigma}, \tilde{\tau}) \equiv \Phi^1(\tilde{\sigma}, \tilde{\tau}) + \tau \Phi^2(\tilde{\sigma}, \tilde{\tau}) \,, \tag{4.11}$$

to be the complex field of the 1+1D σ-model. The boundary conditions are geometrical:

$$\mathcal{Z}(0, \tilde{\tau}) = e^{iv} \mathcal{Z}(2\pi, \tilde{\tau}) \,,$$

where

$$e^{iv} \equiv \mathbf{c}\rho + \mathbf{d} \,.$$

The twist has a number of fixed points z_r ($r = 1, 2, \dots$) that satisfy

$$e^{iv} z_r - z_r \in \mathbb{Z} + \tau \mathbb{Z} \,,$$

i.e., rotation by υ keeps the point on T^2 that is parameterized by z_r invariant. The number of fixed points is as follows:

1. For $\rho = i$ and $\upsilon = \pi/2$, we have $r = 1, 2$ and the fixed points are $z_1 = 0$ and $z_2 = (1 + i)/2$;

2. For $\rho = e^{\pi i/3}$ and $\upsilon = \pi/3$ we have only one fixed point $z_1 = 0$;

3. For $\rho = e^{\pi i/3}$ and $\upsilon = 2\pi/3$ we have 3 fixed points $z_1 = 0$, $z_2 = (\rho + 1)/3$, and $z_3 = 2(\rho + 1)/3$.

Each fixed point z_r defines a different topological sector of the σ-model via the mode expansion

$$\mathcal{Z} = z_r + \sum_{q \in \mathbb{Z} + \frac{\upsilon}{2\pi}} \frac{1}{q} \alpha_q e^{-iq(\tilde{\tau} + \tilde{\sigma})} + \sum_{q \in \mathbb{Z} - \frac{\upsilon}{2\pi}} \frac{1}{q} \widetilde{\alpha}_q e^{-iq(\tilde{\tau} - \tilde{\sigma})} \,.$$

The modes are all fractional, and there are no zero modes here. The number of ground states is therefore the number of topological sectors, labeled by the index r. The result, which is listed above, agrees with the result listed at the end of §4.2.

4.4 Kähler σ-models and the Witten index

The discussion in the previous subsections can be extended to supersymmetric σ-models. We will encounter in §6 the following situation: a selfdual nonlinear 1+1D σ-model with $\mathcal{N} = (2, 2)$ supersymmetry compactified on S^1 with a duality twist augmented by an isometry of the Kähler target space. The duality is mirror symmetry

[37] -[40], and the complex Kähler moduli are at a selfdual value. We denote the target space by Y.

The question is whether the low-energy description (energy scale much lower than the Kaluza-Klein scale of the circle compactification) is a topological theory. We assume that the combination of mirror symmetry and isometry twist preserves some amount of supersymmetry (half of the SUSY generators, generally), and we wish to count the number of vacua, or at least calculate the Witten index.

In our application below, Y will be the Hitchin's moduli space $\mathcal{M}_H$ (to be reviewed in §6.5), which is actually hyper-Kähler, and not just Kähler (and the σ-model therefore starts out with $\mathcal{N} = (4,4)$ supersymmetry in 1+1D). But for the time being it is good to start with a simple prototypical example: $Y = T^2 \times \mathbb{C}$ (which also happens to be hyper-Kähler) where the complexified Kähler class ρ of T^2 is one of the two choices from the end of §4.2. The twist along S^1 is a combination of mirror symmetry and isometry. Mirror symmetry is just the T-duality of T^2 in this context, and is described by an element $\mathbf{t} \in \mathrm{SL}(2,\mathbb{Z})$ which we also pick out from the list at the end of §4.2. We combine it with a rotation of $\mathbb{C}$ by some nonzero angle β. It is not hard to check that none of the fermionic fields of the σ-model have zero-modes, and so it is clear from the discussion above that the low-energy theory is equivalent to geometric quantization of the isometry-invariant subspace of Y, which is $T^2 \times \{0\}$ (where $\{0\}$ stands for the origin of $\mathbb{C}$).

To see this in more detail, let us rederive the results of the previous subsections in this supersymmetric context, using a technique that will be useful in the more complicated case of $\mathcal{M}_H$ later on. Since none of the fermionic fields have zero modes, we can count the number of vacua by calculating the Witten index of the theory. For this purpose, we compactify time on a (Euclidean) circle $0 \le \tilde{\tau} \le 2\pi T$ and calculate the partition function. The fermions have periodic boundary conditions that preserve supersymmetry. We also introduce a topological twist [37] that turns the supersymmetric $\mathcal{N} = (2,2)$ σ-model into either the A-model or the B-model, which we will discuss separately below. Since we are working on a flat worldsheet, the topological twist has no effect on the partition function.

Since the T-duality twist $\mathcal{T}$ commutes with the BRST operator of either A-model or B-model, we can reduce the calculation of the partition function to a trace of the reduction of $(-1)^F\mathcal{T}$ in the finite dimensional Hilbert space of the A-model compactified on S^1 (the $\tilde{\tau}$-direction), where F is the fermion number. The fact that this gives the same result as in §4.3 can be understood as a variant of the Lefschetz–Hopf fixed-point theorem which relates the number of fixed points (counted with multiplicity) of a continuous map on a manifold to the trace of the induced map on cohomology, and can be derived from a topological field theory [41][42][43]. Let us now proceed to the

details.

A-model

For a T^2 target space as in §4.3, the A-model action is:

$$L = \frac{4\pi \operatorname{Im}\rho}{\operatorname{Im}\tau}\int\left(\frac{1}{2}\overline{\partial}_{\overline{z}}\overline{\mathcal{Z}}\partial_z\mathcal{Z} + \frac{1}{2}\overline{\partial}_{\overline{z}}\mathcal{Z}\partial_z\overline{\mathcal{Z}} + i\psi_z\overline{\partial}_{\overline{z}}\chi + i\overline{\psi}_{\overline{z}}\partial_z\overline{\chi}\right)d^2z$$
$$+\frac{2\pi \operatorname{Re}\rho}{\operatorname{Im}\tau}\int\left(\overline{\partial}_{\overline{z}}\overline{\mathcal{Z}}\partial_z\mathcal{Z} - \overline{\partial}_{\overline{z}}\mathcal{Z}\partial_z\overline{\mathcal{Z}}\right)d^2z\,, \tag{4.12}$$

where $z = \tilde{\sigma} + i\tilde{\tau}$, $\mathcal{Z}$ is the same complex coordinate on the target space as in (4.11), $\psi_z, \overline{\psi}_{\overline{z}}, \chi, \overline{\chi}$ are fermionic fields, ρ is the (complex) Kähler modulus of T^2 (taken from the list at the end of §4.2), and τ is the complex structure of T^2, which decouples from the topological theory. The BRST symmetry acts as [37]:

$$\delta\mathcal{Z} = i\epsilon\chi\,,\quad \delta\overline{\mathcal{Z}} = i\epsilon\overline{\chi}\,,\quad \delta\chi = \delta\overline{\chi} = 0\,,\quad \delta\psi_z = -\epsilon\partial_z\overline{\mathcal{Z}}\,,\quad \delta\overline{\psi}_{\overline{z}} = -\epsilon\overline{\partial}_{\overline{z}}\phi\,. \tag{4.13}$$

The Hilbert space of the topological A-model compactified on S^1 (to be understood as the $\tilde{\tau}$ direction, according to the discussion above) is in one-to-one correspondence with the Dolbeault cohomology of T^2. A basis of local BRST-cohomology operators which correspond to these states consists of [37] 1, χ, $\overline{\chi}$, $\overline{\chi}\chi$, which correspond to the following representatives of the Dolbeault cohomology of T^2: 1, dZ, $d\overline{Z}$, $d\overline{Z}\wedge d\overline{Z}$. (Here $Z, \overline{Z}$ are coordinates on T^2 which are in one-to-one correspondence with the σ-model fields $\mathcal{Z}, \overline{\mathcal{Z}}$.)

The T-duality element $\mathcal{T}$ acts on the fermionic fields of the A-model as follows (compare with (4.6)):

$$\chi \to e^{iv}\chi\,,\quad \overline{\chi} \to e^{iv}\overline{\chi}\,,\quad \psi \to e^{-iv}\psi\,,\quad \overline{\psi} \to e^{-iv}\overline{\psi}\,, \tag{4.14}$$

and commutes with the BRST transformation (4.13). The T-duality element $\mathcal{T}$ therefore acts on an A-model operator that corresponds to a (p,q)-Dolbeault cohomology class as multiplication by the phase $e^{i(p+q)v}$. The action depends only on the total degree of the form, as it should, since the A-model is independent of the complex structure of the target space.

Using the state-operator correspondence, we can now determine the action of $\mathcal{T}$ on states, up to a phase. Letting $|1\rangle$ be the state corresponding to the operator 1, the phase is $\langle 1|\mathcal{T}|1\rangle$. The Witten index is then

$$I = \operatorname{tr}\{(-1)^F\mathcal{T}\} = (1 - e^{iv})^2\langle 1|\mathcal{T}|1\rangle.$$

Thus, we get

$$|I| = |1 - e^{-iv}|^2 = 2(1 - \cos v).$$

(And the missing phase is $\langle 1|\mathcal{T}|1\rangle = \pm e^{-iv}$.) This agrees with the results of §4.3.

B-model

The B-model action with T^2 target space is

$$L = \frac{4\pi \operatorname{Im}\rho}{\operatorname{Im}\tau}\int_\Sigma d^2z\Big(\frac{1}{2}\partial_z\mathcal{Z}\partial_{\overline{z}}\overline{\mathcal{Z}} + \frac{1}{2}\partial_z\overline{\mathcal{Z}}\partial_{\overline{z}}\mathcal{Z} + \frac{i}{2}\eta(\partial_z\rho'_{\overline{z}} + \partial_{\overline{z}}\rho'_z) + \frac{i}{2}\theta(\partial_{\overline{z}}\rho'_z - \partial_z\rho'_{\overline{z}})\Big)\,.$$

with the BRST action

$$\delta\mathcal{Z} = 0, \quad \delta\overline{\mathcal{Z}} = i\epsilon\eta, \quad \delta\eta = \delta\theta = 0, \quad \delta\rho' = -\epsilon d\phi\,. \tag{4.15}$$

The BRST-invariant operators are $1, \eta, \theta, \eta\theta$, which correspond to the following elements of $H^p(\wedge^q T^{(1,0)}(T^2))$: $1, d\overline{\mathcal{Z}}, \frac{\partial}{\partial\mathcal{Z}}, d\overline{\mathcal{Z}}\frac{\partial}{\partial\mathcal{Z}}$.

T-duality acts as

$$\eta \to \eta\cos\upsilon + i\theta\sin\upsilon\,, \quad \theta \to i\eta\sin\upsilon + \theta\cos\upsilon\,, \quad \rho'_z \to e^{i\upsilon}\rho'_z\,, \quad \rho'_{\overline{z}} \to e^{-i\upsilon}\rho_{\overline{z}}\,,$$

and we can verify that $\operatorname{tr}\{(-1)^F\mathcal{T}\} = 2 - 2\cos\upsilon$.

5. Analysis for $U(1)$ Super-Yang–Mills

We now study 3+1D Yang–Mills theory with an S-duality twist for the case of a $U(1)$ gauge group. In this case, there is an exact expression for the S-duality kernel, which is well-known, and it is straightforward to find the topological 3D theory associated with the S-duality twist.

5.1 The duality kernel for $U(1)$ Yang–Mills theory

We take pure $U(1)$ Yang–Mills theory with 1-form gauge field A defined on M_3. The S-duality kernel $\mathcal{S}(A, \widetilde{A})$ acts on the wavefunction $\Psi\{A\}$ representing a state so that

$$\widetilde{\Psi}\{A\} \equiv \int [\mathcal{D}\widetilde{A}]\mathcal{S}(A, \widetilde{A})\Psi(\widetilde{A})$$

is the wavefunction of the S-dual state.

For an S-duality transformation we have

$$\tau \to \frac{\mathbf{a}\tau + \mathbf{b}}{\mathbf{c}\tau + \mathbf{d}}\,, \qquad E_i \to \mathbf{a}E_i + \mathbf{b}B_i\,, \quad B_i \to \mathbf{c}E_i + \mathbf{d}B_i\,.$$

The action of S-duality in the quantum theory on an arbitrary manifold was described in [44]. A closed expression for the S-duality kernel appears in [38][45][13]:

$$\mathcal{S}_A(A, \widetilde{A}) = \exp\left\{\frac{i}{4\pi\mathbf{c}}\int(\mathbf{d}A\wedge dA - 2\widetilde{A}\wedge dA + \mathbf{a}\widetilde{A}\wedge d\widetilde{A})\right\}. \tag{5.1}$$

This is determined by requiring the operator equations

$$\widetilde{E}_i \mathcal{S}_A = \mathcal{S}_A(\mathbf{a}E_i + \mathbf{b}B_i), \qquad \widetilde{B}_i \mathcal{S}_A = \mathcal{S}_A(\mathbf{c}E_i + \mathbf{d}B_i).$$

Here we can take $E_i \equiv -2\pi i\delta/\delta A_i$.

Now set $A = \widetilde{A}$ (up to a gauge transformation) in (5.1). We get

$$\mathcal{I}(A) \equiv \frac{\mathbf{a}+\mathbf{d}-2}{4\pi\mathbf{c}} \int A \wedge dA\,.$$

This is a Chern–Simons theory at level $k \equiv (\mathbf{a}+\mathbf{d}-2)/\mathbf{c}$. For a generic $\mathrm{SL}(2,\mathbb{Z})$ element $\mathbf{s} = \begin{pmatrix} \mathbf{a} & \mathbf{b} \\ \mathbf{c} & \mathbf{d} \end{pmatrix}$ this is not an integer, but for the special values $\mathbf{s} = \mathbf{s}', -\mathbf{s}', \mathbf{s}'', -\mathbf{s}''$ we get integral levels $k = -2, 2, -1, 3$, respectively.

5.2 Low-energy limit of an $\mathrm{SL}(2,\mathbb{Z})$-twisted compactification

Now, let us compare the Chern–Simons action that we obtain from the diagonal $\mathcal{S}_A(A, A)$ to the action that we obtain from compactifying $U(1)$ Yang–Mills theory on S^1 of radius R with an $\mathbf{s}$-twist, in the limit $R \to 0$, as in §2.1.

Let us describe the full action in detail. We assume that the $\mathbf{s}$-twist is at $x_3 = 0 \simeq 2\pi R$. The Yang–Mills field $A(x_0, x_1, x_2, x_3)$ is defined in the range $0 \le x_3 \le 2\pi R$ *without* imposing periodic boundary conditions. The Yang–Mills coupling constant is either $\tau = i$ or $\tau = e^{\pi i/3}$, according to whether $\mathbf{s} = \mathbf{s}'$ or $\mathbf{s} = \pm\mathbf{s}''$. As is customary, we set $\tau \equiv \tau_1 + i\tau_2$. We also denote

$$A' \equiv A(x_0, x_1, x_2, x_3 = 0)\,, \qquad A'' \equiv A(x_0, x_1, x_2, x_3 = 2\pi R)\,.$$

The full action is

$$\mathcal{I} = \mathcal{I}_{YM} + \mathcal{I}_X\,,$$

where $\mathcal{I}_{YM}$ is the bulk Yang–Mills action

$$\mathcal{I}_{YM} \equiv \int_{x_0,x_1,x_2} \int_{x_3=0}^{2\pi R} \left(\frac{1}{2g_{\mathrm{YM}}^2} F \wedge {}^*F + \frac{\theta}{4\pi^2} F \wedge F \right)\,,$$

and $\mathcal{I}_X$ consists of "boundary terms"

$$\mathcal{I}_X \equiv \frac{1}{4\pi} \int_{x_0,x_1,x_2} \omega\,,$$

where

$$\omega \equiv \frac{1}{\mathbf{c}} \left(\mathbf{d}A' \wedge dA'' - 2A' \wedge dA'' + \mathbf{a}A'' \wedge dA'' \right)$$

is the integral of a gauge-invariant expression:

$$d\omega = \frac{1}{\mathbf{c}}(\mathbf{d}F' \wedge F'' - 2F' \wedge F'' + \mathbf{a}F'' \wedge F'') .$$

The equations of motion are Maxwell's equations in the bulk, but with boundary conditions:

$$F'' = (\mathbf{c}\tau_1 + \mathbf{d})\, F' + \mathbf{c}\tau_2\, {}^*F' , \qquad F' = -(\mathbf{c}\tau_1 - \mathbf{a})\, F'' - \mathbf{c}\tau_2\, {}^*F'' .$$

These two conditions are equivalent for selfdual values of τ and corresponding $\mathbf{s}$.

Define the complex-valued 2-forms

$$F'_\pm \equiv {}^*F' \pm iF' , \qquad F''_\pm \equiv {}^*F'' \pm iF'' .$$

Then, the boundary conditions can be written as

$$F''_+ = (\mathbf{c}\tau + \mathbf{d})F'_+ , \qquad F''_- = (\mathbf{c}\overline{\tau} + \mathbf{d})F'_- . \tag{5.2}$$

As noted earlier, $|\mathbf{c}\tau + \mathbf{d}| = 1$ so that we can write $\mathbf{c}\tau + \mathbf{d} = e^{2\pi i q}$ with

$$q = \begin{cases} \frac{1}{4} \text{ for } \tau = i \text{ and } \mathbf{s} = \mathbf{s}', \\ \frac{1}{6} \text{ for } \tau = e^{\pi i/3} \text{ and } \mathbf{s} = \mathbf{s}'', \\ \frac{2}{3} \text{ for } \tau = e^{\pi i/3} \text{ and } \mathbf{s} = -\mathbf{s}''. \end{cases}$$

From the boundary conditions (5.2) we find the Fourier mode decomposition

$$F_\pm = \sum_{j\in\mathbb{Z}} e^{\frac{i(j+q)x_3}{R}} f^{(+)}_{j+q}(x_0, x_1, x_2) .$$

Because $j+q$ is never zero, we see that the fields $f^{(+)}_{j+q}$ are massive in 2+1D with masses given by $|j+q|/R$. The classical analysis, however, cannot tell us the multiplicity of the vacuum. But since the low-energy description is a Chern–Simons theory at level $k \equiv (\mathbf{a} + \mathbf{d} - 2)/\mathbf{c}$ we expect to get a multiplicity of k^h vacua when formulated on a compact genus-h Riemann surface $\mathcal{C}_h$.

5.3 Supersymmetry

Now let us extend the discussion to a free vector multiplet of $\mathcal{N} = 4$ SYM. The extra fields are free scalars and fermions. We need to impose the boundary conditions (2.9) combined with the $\mathbf{s}$-twist. The action of $\mathbf{s}$ on the 6 scalar fields of the vector multiplet is trivial, since it commutes with the $SO(6)$ R-symmetry. One might consider the possibility of $\mathbf{s}$ acting as an overall $(-)$ sign, which corresponds to the nontrivial

element in the center of $SO(6)$, but this is a matter of definition, and we can always absorb it in the R-symmetry twist γ. We then get 2+1D scalar Klauza-Klein modes with masses $(j + (\varphi_a + \varphi_b)/2\pi)/R$ $(1 \le a < b \le 4)$, where $j \in \mathbb{Z}$ and φ_a are as in (2.7).

Now consider the free fermions of the $\mathcal{N} = 4$ vector multiplet. By (2.6) and (2.9), their boundary conditions are

$$\psi^{\alpha a}(x_3 = 2\pi R) = e^{\frac{i}{2}\upsilon + i\varphi_a}\psi^{\alpha a}(x_3 = 0)\,, \qquad \overline{\psi}^{\dot\alpha}_a(x_3 = 2\pi R) = e^{-\frac{i}{2}\upsilon - i\varphi_a}\overline{\psi}^{\dot\alpha}_a(x_3 = 0)\,.$$

This gives 2+1D fermionic Klauza-Klein modes with masses $(j + (\varphi_a + \frac{1}{2}\upsilon)/2\pi)/R$ $(1 \le a \le 4)$.

For a generic choice of γ (i.e., generic φ_a) there are neither fermionic nor bosonic zero modes. This is also the case for the $\mathcal{N} = 6$ supersymmetric γ in (2.11). For the $\mathcal{N} = 4$ supersymmetric choices of γ in (2.12) there are no zero modes unless the phase φ_4 is chosen so that $e^{i(\frac{1}{2}\upsilon + \varphi_4)} = 1$. In that case the subgroup of $SU(4)_R$ that commutes with γ is $(SU(2) \times SU(2) \times U(1))/\mathbb{Z}_2$. The surviving supercharges transform in the representation $(\mathbf{2}, \mathbf{1})_{+1} \oplus (\mathbf{2}, \mathbf{1})_{-1}$ so the $U(1)$ factor and the leftmost $SU(2)$ factor can be considered an R-symmetry of the resulting theory, while the right $SU(2)$ factor is a flavor symmetry. The low-energy theory comprises of 4 massless scalar fields in the representation $(\mathbf{2}, \mathbf{2})_0$ of the unbroken R-symmetry, and 4 massless fermions in the representation $(\mathbf{1}, \mathbf{2})_{+1} \oplus (\mathbf{1}, \mathbf{2})_{-1}$. These combine to a 2+1D hypermultiplet. The moduli space is $\mathbb{R}^4$.

So far we discussed the physical low-energy theory. Now, let us discuss the action defined by (3.4). The duality kernel is given by

$$\mathcal{S}(V, \widetilde{V}) = \mathcal{S}_A(A, \widetilde{A})\delta(\widetilde{\Phi} - \Phi^\gamma)\delta(\widetilde{\psi} - e^{\frac{i}{2}\upsilon}\psi^\gamma)\,,$$

where $\mathcal{S}_A(A, \widetilde{A})$ is given by (5.1). Setting $\widetilde{V} = V$ we get, up to an infinite normalization factor,

$$\mathcal{S}(V, V) = e^{\frac{ik}{4\pi}\int A \wedge dA}\delta(\Phi)\delta(\psi). \tag{5.3}$$

The normalization factor is, formally, a product of the determinants (one determinant for each spacetime point x), since

$$\delta(\Phi(x) - \Phi(x)^\gamma)\delta(\psi(x) - e^{\frac{i}{2}\upsilon}\psi(x)^\gamma) = \delta(\Phi(x))\delta(\psi(x))\frac{\prod_a(1 - e^{i(\varphi_a + \frac{1}{2}\upsilon)})}{\prod_{a<b}(1 - e^{i(\varphi_a + \varphi_b)})}\,.$$

If the constant factor on the right is well-defined, nonzero, and finite (i.e., in the absence of fermionic and bosonic zero modes) the resulting action $\mathcal{S}(V, V)$ is indeed topological, after regularization.

Note that at low-energy, in the topological theory, supersymmetry now acts in a trivial way: all the supersymmetry generators are identically zero. This is because by (5.3) we have $\Phi = 0$ and $\psi = 0$, and the equations of motion of the Chern–Simons theory also set $F = 0$. The vanishing of the SUSY generators immediately implies that the Hamiltonian is identically zero (since the Hamiltonian is part of the supersymmetry algebra), which is consistent with the topological nature of the low-energy theory.

6. The nonabelian case

We now turn to the nonabelian case. Our setting is $\mathcal{N} = 4$ $SU(n)$ SYM compactified on S^1 of radius R with an R-symmetry twist γ and an $\mathrm{SL}(2,\mathbb{Z})$-duality twist $\mathbf{s}$ at a point on the circle.[5] We have argued that the low-energy limit, $R \to 0$, is described by a 2+1D topological field theory. We ask: *what is that field theory?*

In §6.1 we present our conjecture: the low-energy limit can be described by a Chern–Simons theory at a level that is determined by the twist. We then test this conjecture in §6.4 by calculating the Witten index of the theory compactified (in an appropriate way that preserves some supersymmetry) on a Riemann surface, and we compare the result to the number of vacua of Chern–Simons theory on that Riemann surface. We now proceed to the details.

6.1 A conjecture

Our conjecture is as follows. For the values of $n, \tau, \mathbf{s}, \upsilon$ listed below, the low-energy limit of $\mathcal{N} = 4$ SYM with gauge group $SU(n)$ and complex coupling constant τ, compactified on S^1 with an $\mathrm{SL}(2,\mathbb{Z})$-twist $\mathbf{s}$ and R-symmetry twist γ (determined by υ) as in (2.11), is described by a (three-dimensional) pure Chern–Simons theory with the same gauge group $SU(n)$ and at level k that is given by:

- for $\tau = i$, $\upsilon = \frac{\pi}{2}$, $\mathbf{s} = \mathbf{s}' \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$, and $n = 1, 2, 3$, we have $k = -2$;
- for $\tau = e^{\pi i/3}$, $\upsilon = \frac{\pi}{3}$, $\mathbf{s} = \mathbf{s}'' \equiv \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix}$, and $n = 1, 2, 3, 4, 5$, we have $k = -1$;
- for $\tau = e^{\pi i/3}$, $\upsilon = \frac{4\pi}{3}$, $\mathbf{s} = -\mathbf{s}'' = \begin{pmatrix} -1 & 1 \\ -1 & 0 \end{pmatrix}$, and $n = 1, 2$, we have $k = 3$.

[5] The $SU(n)$ theory is not selfdual under the full $\mathrm{SL}(2,\mathbb{Z})$ group, but rather only under a subgroup known as $\Gamma_0(n)$. This is because the dual group of $SU(n)$ is its adjoint form $SU(n)/\mathbb{Z}_n$. The difference has to do with allowed electric and magnetic fluxes, which we will address in §6.6. For the time being, we will ignore this subtlety.

Supersymmetry is realized trivially (all generators are zero). The levels k are conjectured by extension from the $U(1)$ case discussed in §5.1. The restrictions on the rank n are in order to eliminate zero-modes of scalar fields, as discussed in §2.3. The negative values of k for the first two cases in the list can, of course, be flipped to positive values with the help of a parity transformation.

The conjecture implies that the expectation value of a large smooth Wilson loop can be calculated from Chern–Simons theory. In Euclidean signature, let $0 \leq x_3 < 2\pi R$ be a periodic coordinate on S^1, and let $\mathbf{C} \subset \mathbb{R}^3$ be a loop at a constant x_3 (and here $\mathbb{R}^3$ represents the remaining three dimensions of the problem). We assume that the curvature of $\mathbf{C}$ is small compared to R^{-1} and that the loop is not self-intersecting or "close" to being self-intersecting. (More precisely, we assume that the intersection of $\mathbf{C}$ with any ball in $\mathbb{R}^3$ of radius of the order of R or less is topologically connected.) The expectation value $\langle W(\mathbf{C})\rangle$ of a Wilson loop $W(\mathbf{C})$ is then given, by conjecture, by a similar expectation value $\langle W(\mathbf{C})\rangle$ in the corresponding three-dimensional Chern–Simons theory. It can therefore be calculated using the techniques developed in [25].

The restriction on the curvature of the loop can presumably be dropped if we supersymmetrize the loop, as in [46][47]. Since the scalars and fermions are set to zero at low-energy, by our conjecture, the supersymmetrization should have no effect on the Chern–Simons side.

6.2 Relations among the Chern–Simons levels

The three cases corresponding to the twists $\mathbf{s} = \mathbf{s}', \mathbf{s}'', -\mathbf{s}''$ are related, and were it not for the different γ-twists, a proof of the conjecture for any one of them would have implied the rest. To see this, set

$$T \equiv \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad S \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Then

$$\mathbf{s}' = S\,, \qquad \mathbf{s}'' = TS\,, \qquad -\mathbf{s}'' = TS^{-1}\,.$$

The action of T is simple to describe. It multiplies the wavefunction of 3+1D SYM by the level $k = 1$ Chern–Simons phase, as in (3.2). Furthermore, if the kernel for S is $\mathcal{S}(V, \widetilde{V})$, in the notation of §3.2, then the kernel for S^{-1} is $\mathcal{S}(\widetilde{V}, V)^*$, since $\hat{\mathcal{S}}$ is a unitary operator. It follows that if the diagonal of the kernel for $\mathbf{s}'$ corresponds to Chern–Simons theory at level k, then $\mathbf{s}''$ is described by level $(k+1)$ and $-\mathbf{s}''$ by level $(1-k)$ (which happens to be true from the list of §6.1). However, since the R-symmetry twists are different in the three cases, we do not know how to prove this relation definitively.

6.3 Compactification on a Riemann surface $\mathcal{C}_h$

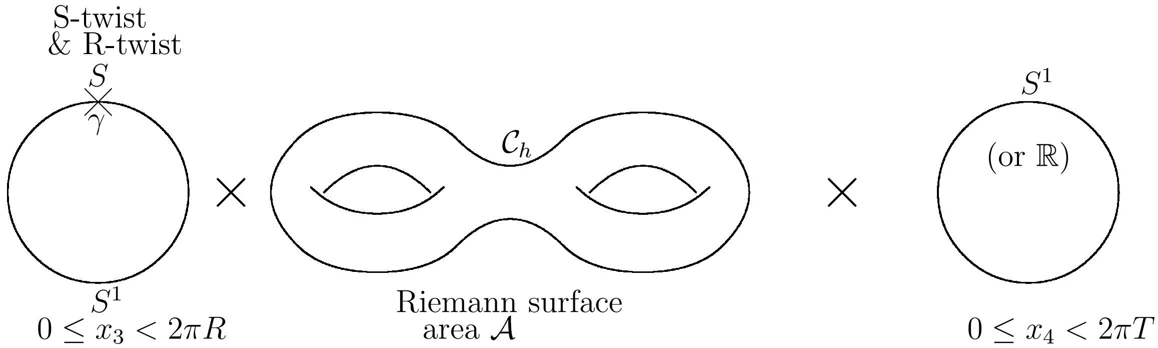

Figure 1: Our setting is $\mathcal{N} = 4$ $SU(n)$ SYM compactified on an S^1 with an R-symmetry and S-duality twist times a Riemann surface $\mathcal{C}_h$ of genus h ($h = 2$ in the picture). The remaining dimension is also compactified on another S^1. The R-symmetry bundle is nontrivial over $\mathcal{C}_h$ so as to preserve half of the supersymmetry.

In order to explore the conjecture presented in §6.1 we wish to find a topological quantity that can be computed in $\mathcal{N} = 4$ SYM, using what is already known about the action of S-duality, and then compare the result to what our conjecture predicts in terms of Chern–Simons theory. As a first step, we compactify the theory on a Riemann surface $\mathcal{C}_h$ of genus h. In other words, we consider the theory on $X = S^1_R \times \mathcal{C}_h \times \mathbb{R}$, where the subscript R refers to the radius of the circle, with the γ and $\mathbf{s}$ twists setting the boundary conditions along S^1_R.

We also wish to preserve some amount of supersymmetry, so that Witten-index techniques could be applicable. We can do this by turning on an appropriate (topologically nontrivial) background gauge field along $\mathcal{C}_h$ for the unbroken R-symmetry—an operation known as "twisting" [37, 48], which we will briefly review.

For this additional twisting we are only allowed to use the unbroken subgroup of the R-symmetry group. The already present R-twist of (2.11) breaks the R-symmetry group of $\mathcal{N} = 4$ SYM down to $U(3) \subset SU(4)_R$, under which the 6 supercharges transform as the sum of the fundamental and anti-fundamental representations $\mathbf{3} + \overline{\mathbf{3}}$. The supercharges also transform as a spinor (left-moving plus right-moving) in the two directions of $\mathcal{C}_h$, which means that one component transforms as a section of the $SO(2)$ bundle associated with the phase of the square-root of the canonical line bundle $\mathcal{K}$ of $\mathcal{C}_h$ and the other component transforms as a section of the opposite bundle (the one associated with the anti-canonical bundle $\overline{\mathcal{K}}$). For genus $h \neq 1$ these are nontrivial bundles,

and there are therefore no covariantly constant spinors on $\mathcal{C}_h$, and supersymmetry is completely broken.

The procedure of twisting restores supersymmetry by adding a background $SU(4)_R$ gauge field that is proportional to the spin connection of $\mathcal{C}_h$. This modifies the covariant derivative of the fermions and scalars that are charged under $SU(4)_R$. The spin connection of $\mathcal{C}_h$ can be thought of as a gauge field for the group of rotations $SO(2)$ of the fibers of the tangent-bundle of $\mathcal{C}_h$. To specify the topological twist we need to specify an element $\mathbf{T}$ in the (R-symmetry) Lie algebra $\mathfrak{su}(4)$. Denoting by ω_j ($j = z, \overline{z}$) the components of the spin connection on $\mathcal{C}_h$, the covariant derivative of a left-moving fermion in the $\mathcal{C}_h$ direction j is then $D_j = \partial_j - \frac{1}{2}\omega_j - \omega_j \mathbf{T}$. Here $\mathbf{T}$ acts on the R-symmetry indices of the field, and we assume that it commutes with the R-symmetry twist in (2.11): $\mathbf{T} \in \mathfrak{u}(3) \subset \mathfrak{su}(4)$. In the basis that corresponds to (2.7) we therefore take

$$\mathbf{T} \equiv \begin{pmatrix} \varrho_1 & & & \\ & \varrho_2 & & \\ & & \varrho_3 & \\ & & & -\sum_1^3 \varrho_i \end{pmatrix} \in \mathfrak{u}(3) \subset \mathfrak{su}(4)_R \,. \tag{6.1}$$

After the topological twist (and contraction with the zweibein if necessary), the scalars and fermions turn into sections of generally nontrivial line bundles over $\mathcal{C}_h$ which are certain powers of $\mathcal{K}$ or $\overline{\mathcal{K}}$. The supercharges are also sections of such line bundles, and the number of conserved supersymmetries is the number of supercharges that transform in the trivial bundle [37].

The supercharges that transform in the $\mathbf{4}$ of $SU(4)_R$ are also left-moving spinors under the 3+1D Lorentz group, and they break up into two components: a left-mover on $\mathcal{C}_h$ which is also a left-mover on the remaining two dimensions $S^1 \times \mathbb{R}$, and a right-mover on $\mathcal{C}_h$ which is also a right-mover on the remaining two dimensions $S^1 \times \mathbb{R}$. Altogether, therefore, the supercharges transform as a section of the following vector bundle [where the subscript indicates whether it is a left-mover (+) or right-mover (−) on $S^1 \times \mathbb{R}$]:

$$\left[\mathcal{K}^{\frac{1}{2}-\sum_1^3 \varrho_i} \oplus \bigoplus_{i=1}^{3} \mathcal{K}^{\frac{1}{2}+\varrho_i}\right]_+ \oplus \left[\mathcal{K}^{\frac{1}{2}+\sum_1^3 \varrho_i} \oplus \bigoplus_{i=1}^{3} \mathcal{K}^{\frac{1}{2}-\varrho_i}\right]_- .$$

At the same time, scalar fields transform as sections of

$$\mathcal{K}^{\varrho_1+\varrho_2} \oplus \mathcal{K}^{\varrho_1+\varrho_3} \oplus \mathcal{K}^{\varrho_2+\varrho_3} \,, \tag{6.2}$$

and their complex conjugates, of course, transform as sections of

$$\overline{\mathcal{K}}^{\varrho_1+\varrho_2} \oplus \overline{\mathcal{K}}^{\varrho_1+\varrho_3} \oplus \overline{\mathcal{K}}^{\varrho_2+\varrho_3} \,. \tag{6.3}$$

The maximum number of supersymmetry generators that can be preserved is 4. For this we take $\varrho_1 = \varrho_2 = -\varrho_3 = \frac{1}{2}$, i.e.,

$$\mathbf{T} \equiv \begin{pmatrix} \frac{1}{2} & & & \\ & \frac{1}{2} & & \\ & & -\frac{1}{2} & \\ & & & -\frac{1}{2} \end{pmatrix} \in \mathfrak{u}(3) \subset \mathfrak{su}(4)_R \,. \tag{6.4}$$

This is the A-twist discussed in [6][21]. We see from (6.2)-(6.3) that the 6 real scalars of $\mathcal{N} = 4$ SYM have turned into 4 scalars and a 1-form on $\mathcal{C}_h$.

We thus end up with the following setting: $\mathcal{N} = 4$ SYM compactified on $S^1_R \times \mathcal{C}_h$ with an R-symmetry twist γ and an $\mathrm{SL}(2,\mathbb{Z})$-twist $\mathbf{s}$ along S^1_R as throughout this paper, and with an additional A-twist along $\mathcal{C}_h$. We wish to find the Witten index, i.e., the number of supersymmetric vacua counted with $(\pm)$ signs according to their fermion numbers.

6.4 The Witten Index

The Witten index is generally independent of continuous parameters, and as is standard in the computation of a Witten index, when we identify a useful parameter that can be taken to different extreme values we get two opposite limits in which it is interesting to perform the calculation. In our case, one limit is that the Riemann surface $\mathcal{C}_h$ is much larger than the circle S^1_R. We refer to it as "Limit (i)." In this case, we first reduce to Chern–Simons theory on $\mathcal{C}_h \times \mathbb{R}$, according to our conjecture in §6.1, and viewing $\mathbb{R}$ as time direction, the Witten index is just the dimension of the Hilbert space of Chern–Simons theory.

The method for calculating the dimension $d_h(n,k)$ of the Hilbert space of $SU(n)$ Chern–Simons theory at level k on a Riemann surface of genus h was outlined in [25]. The Hilbert space can be obtained by geometric quantization of the moduli space $\mathcal{M}_{\text{fc}}$ of flat $SU(n)$ connections on $\mathcal{C}_h$ with a symplectic form that is k times the Kähler 2-form of $\mathcal{M}_{\text{fc}}$, which is determined by the complex structure of $\mathcal{C}_h$. Explicit expressions for $n = 2$ can be found in [49]. However, as we will see later, we need to modify these equations to include a nonzero magnetic flux through $\mathcal{C}_h$, and we present the calculation in Appendix A.

The opposite limit is to take $\mathcal{C}_h$ to be much smaller than S^1_R. We refer to it as "Limit (ii)." We can then first reduce $\mathcal{N} = 4$ SYM on $\mathcal{C}_h$. This is precisely the setting studied in [6][7][21]. With a nonzero magnetic flux on $\mathcal{C}_h$, the resulting low-energy description is a 1+1D σ-model with a smooth hyper-Kähler target space that can be identified with Hitchin's moduli space $\mathcal{M}_H$. (We will review Hitchin's space in §6.5.)

The magnetic flux is required to make the associated Hitchin space $\mathcal{M}_H$ nonsingular. S-duality, according to [6][7], reduces to T-duality of the σ-model.

To compute the Witten index, we compactify the $\mathbb{R}$ direction on a circle of radius T. (The resulting setting is depicted in Figure 1.) We take periodic boundary conditions along the S^1_T direction for all the σ-model fields, and calculate the index in the limit $R \ll T$. In this limit it is convenient to switch the roles of time and space and think of S^1_R as (Euclidean) time. The Witten index is then given by the trace of the T-duality operator $\mathcal{T}(\mathbf{s})$, which is the reduction of the $\mathrm{SL}(2,\mathbb{Z})$ twist $\mathbf{s}$ to the Hilbert space of the σ-model compactified on S^1, times the R-symmetry operator γ, treated as an operator in the same Hilbert space (we hope the reader will forgive this slight abuse of notation):

$$I = \mathrm{tr}_0\{(-1)^F \mathcal{T}(\mathbf{s})\gamma\}\,. \tag{6.5}$$

Here F is the fermion number, and tr_0 denotes the restriction of the trace to the ground states.

The ground states form a finite dimensional Hilbert space which can be identified with the cohomology of the target space $\mathcal{M}_H$. In fact, since the two dimensional space on which the σ-model is defined is flat, we can topologically twist the σ-model to get an A-model or B-model [37] with the same target space. There is a particular complex structure on $\mathcal{M}_H$ for which the A- and B-models are invariant under S-duality (called "complex structure I" in [21]). But in any case, since $\mathcal{T}(\mathbf{s})\gamma$ preserves supersymmetry it commutes with the BRST charge, and hence acts on the finite dimensional Hilbert space of the topologically twisted theory. This Hilbert space is identified with the (de Rham or Dolbeault) cohomology of $\mathcal{M}_H$, and in order to complete the computation of the Witten index we need to know how γ and $\mathcal{T}(\mathbf{s})$ act on the cohomology.

6.5 Review of Hitchin's space

It is now time to review some relevant facts about Hitchin's moduli space $\mathcal{M}_H = \mathcal{M}_H(\mathcal{C}_h, G)$ associated with a Riemann surface $\mathcal{C}_h$ and a gauge group G. What follows is a list of facts that are relevant to our discussion, collected from [6, 21, 50, 51].

The Hitchin moduli space is the moduli space of solutions to the following differential equations:

$$F_{z\overline{z}} = [\phi_z, \overline{\phi}_{\overline{z}}]\,, \qquad D_z\overline{\phi}_{\overline{z}} = D_{\overline{z}}\phi_z = 0\,, \tag{6.6}$$

where solutions that are equivalent up to a gauge transformation are identified in the moduli space. Here $F_{z\overline{z}}$ is the field strength of a gauge field with gauge group G on $\mathcal{C}_h$, $\phi_z dz$ is a $(1,0)$-form which takes values in the complexified Lie algebra of G, $\overline{\phi}_{\overline{z}} d\overline{z}$ is its complex conjugate, and $D_z \equiv \partial_z - A_z$ and $D_{\overline{z}} \equiv \partial_{\overline{z}} - A_{\overline{z}}$ are the $(1,0)$ and $(0,1)$ parts

of the covariant derivative. (Here, $A_{\overline{z}} = -A_z^\dagger$.) We focus on the case with $G = SU(2)$ and assume that the genus h of $\mathcal{C}_h$ is greater than 1.

The moduli space $\mathcal{M}_H$ in general contains singularities; these points correspond to reducible solutions of Hitchin's equation. In physical terms, this means that the low energy description of $\mathcal{N} = 4$ SYM in terms of σ-model breaks down at these points due to the presence of massless modes associated with the residual gauge theory. The problem was circumvented in [6] by turning on a nontrivial 't Hooft magnetic flux through $\mathcal{C}_h$. In fact, one of the main results of [50] was that the moduli space $\mathcal{M}_H$ becomes a smooth manifold of dimension $12h - 12$ in this case. Therefore, from now on, we will concentrate on the moduli space of solutions with magnetic flux turned on.

6.5.1 Hitchin's fibration

The crucial point in understanding the T-duality of the σ-model with target space $\mathcal{M}_H$ is what is called *Hitchin's first fibration* in [21]. In this fibration, the base space B is simply parameterized by the gauge-invariant polynomials in ϕ_z; for $G = SU(2)$, this is just $b_{zz} = \operatorname{tr}\phi_z^2$, which is holomorphic due to Hitchin's equations (6.6), and hence belongs to $H^0(\mathcal{C}_h, \mathcal{K}^2) \approx \mathbb{C}^{3h-3}$, where $\mathcal{K}$ is the canonical bundle on $\mathcal{C}_h$. The projection map of the fibration simply sends the pair $(A, \phi_z dz)$ to $b_{zz} = \operatorname{tr}\phi_z^2$.

At a generic point on the base space $H^0(\mathcal{C}_h, \mathcal{K}^2)$, the holomorphic differential b_{zz} has simple zeroes on $\mathcal{C}_h$. To obtain the fiber space over this point, one first constructs a double cover $\hat{\mathcal{C}}_h$ of $\mathcal{C}_h$, determined by the two-valued differential $\sqrt{b_{zz}}$. It is shown in [50] (see also [6]) that the fiber over b_{zz} is then the Prym variety of the double cover $\hat{\mathcal{C}}_h$. (Roughly speaking, this is the space of allowed values of $U(1)$ holonomies, where the $U(1) \subset SU(2)$ is determined by the values of ϕ_z away from the branch points of the double cover.) In particular, the fiber is a complex torus with (complex) dimension $3h - 3$.

6.5.2 The most singular fiber

While the generic fiber of Hitchin's fibration is T^{6h-6}, there are singular fibers as well at special values of the holomorphic quadratic differential $b_{zz} = \operatorname{tr}(\phi_z^2)$. The most singular fiber is over the base point where the quadratic differential is identically zero: $b_{zz} = 0$. This implies that up to an $SU(2)$ gauge transformation ϕ_z takes the form

$$\phi_z = \begin{pmatrix} 0 & \alpha_z \\ 0 & 0 \end{pmatrix} . \tag{6.7}$$

A special case is when $\phi_z = 0$ identically. The solution to Hitchin's equations then reduces to finding a flat connection. Thus $\mathcal{M}_{\mathrm{fc}}$, the moduli space of flat connections

(for a given magnetic flux), is a subset of the fiber over $b_{zz} = 0$. The space $\mathcal{M}_{\rm fc}$ is of dimension $6h - 6$, so it has the same dimension as the fiber.

If ϕ_z is not identically zero, then from (6.6) and (6.7) it is easy to check that the gauge field must take the form:

$$A_{\overline{z}} = \begin{pmatrix} a_{\overline{z}} & c_{\overline{z}} \\ 0 & -a_{\overline{z}} \end{pmatrix} , \tag{6.8}$$

where

$$a_{\overline{z}} = -\frac{1}{2}\partial_{\overline{z}} \log \alpha_z \, ,$$

and $c_{\overline{z}}$ is arbitrary. The equation $F_{z\overline{z}} = [\phi_z, \overline{\phi}_{\overline{z}}]$ implies that $c^*_{\overline{z}}/\alpha_z$ is holomorphic, and that $\partial_z a_{\overline{z}} - \partial_{\overline{z}} a_z = |\alpha_z|^2 + |c_{\overline{z}}|^2$.

A special case of this is when $c_{\overline{z}} = 0$ identically. In what follows, we will only need the case of genus $h = 2$ and with one unit of magnetic flux on $\mathcal{C}_2$. It can then be shown (see §7 of [50]) that α_z has a single simple zero on $\mathcal{C}_2$, and the location of this zero uniquely determines α_z up to a gauge transformation in $SO(3) = SU(2)/\mathbb{Z}_2$. ($\alpha_z$ is not locally holomorphic, but can be written as a product of a section of a holomorphic line bundle times a nonzero function.) There is an extra complication here due to the center $\mathbb{Z}_2$ of the gauge group. If we identify solutions up to *any* gauge transformation in $SO(3)$, including large gauge transformations, then the space of solutions with $c_{\overline{z}} = 0$ can be identified with a copy of $\mathcal{C}_h$. (The map from the moduli space of solutions with $c_{\overline{z}} = 0$ to $\mathcal{C}_h$ is given by the location of the zero of α_z.) But if we identify solutions only up to gauge transformations in $SU(2)$, we have to take into account the existence of $2^{2h} = 16$ classes of large gauge transformations. Each class is characterized by a map $\pi_1(\mathcal{C}_2) \to \mathbb{Z}_2$ which adds $(\pm)$ signs to the holonomies of the abelian gauge field $a_{\overline{z}} d\overline{z} + a_z dz$ along one-cycles of $\mathcal{C}_2$. In this case the space of solutions is a 16-fold cover of $\mathcal{C}_2$, which is a Riemann surface of genus 17. This extra complication will not be important for us, as we will need only the sector with zero electric flux along one-cycles of $\mathcal{C}_2$, and so for all intents and purposes of this paper, the space of solutions with $c_{\overline{z}} = 0$ is identified with $\mathcal{C}_2$. (For more details see §7 of [50].)

To obtain more information on the singular fiber, consider the (real) "Morse function," introduced by Hitchin, on $\mathcal{M}_H$:

$$\mu \equiv 2 \int {\rm tr}(\phi_z \overline{\phi}_{\overline{z}}) d^2 z.$$

The integral is over $\mathcal{C}_h$. Its minimum is $\mu = 0$ and the minimum locus $\mu^{-1}(0)$ is identified with the subspace $\mathcal{M}_{\rm fc}$ of the singular fiber. For $h = 2$, the range of μ on the singular fiber is $0 \leq \mu \leq \frac{\pi}{2}$, and the maximal value $\frac{\pi}{2}$ is attained on the subspace of solutions

with $c_{\overline{z}} = 0$. We will therefore refer to this subspace as $\mu^{-1}(\frac{\pi}{2})$, and as we have just seen, it is isomorphic to a copy of $\mathcal{C}_2$.

Thus, the part of the singular fiber that is not contained in $\mathcal{M}_{\text{fc}}$ is the subset on which μ takes nonzero values, i.e., $\mu^{-1}((0, \frac{\pi}{2}])$, where $(0, \frac{\pi}{2}]$ is the set of values $0 < \mu \leq \frac{\pi}{2}$. For genus $h = 2$, the set $\mu^{-1}((0, \frac{\pi}{2}])$ is an open manifold of (real) dimension 6, while $\mathcal{M}_{\text{fc}} = \mu^{-1}(0)$ is a closed manifold, also of (real) dimension 6. The boundary of $\mu^{-1}((0, \frac{\pi}{2}])$ is a Riemann surface that is a subset of $\mathcal{M}_{\text{fc}}$ and is isomorphic to the Riemann surface $\mu^{-1}(\frac{\pi}{2})$ (i.e., is $\mathcal{C}_2$ if we ignore the 2^4 multiplicity). To see this note that the boundary of $\mu^{-1}((0, \frac{\pi}{2}])$ is obtained by setting $\alpha_z = 0$ in (6.7), but keeping the upper triangular form (6.8) for the gauge field. Define the complex conjugate field $c_z = c^*_{\overline{z}}$. Then, Hitchin's equations (6.6) reduce to

$$a_{\overline{z}} = -\frac{1}{2}\partial_{\overline{z}} \log c_z \,, \qquad \partial_z a_{\overline{z}} - \partial_{\overline{z}} a_z = |c_z|^2 \,.$$

But these are the same equations that $a_{\overline{z}}$ and α_z satisfy on $\mu^{-1}(\frac{\pi}{2})$, only that the role of α_z is played by c_z. Thus, the boundary of $\mu^{-1}((0, \frac{\pi}{2}])$, as $\mu \to 0$, is isomorphic to $\mu^{-1}(\frac{\pi}{2})$.

6.5.3 Cohomology

In what follows, we will also need some facts about the cohomology $H^*(\mathcal{M}_H)$. We will restrict to the case of gauge group $SU(2)$ and genus $h = 2$.

The Poincaré polynomial

$$P(t) \equiv \sum_i \dim H^i(\mathcal{M}_H) t^i \tag{6.9}$$

was calculated in [50], and is given by

$$P(t) = 1 + t^2 + 4t^3 + t^4 + t^6 + t^4(1 + 34t + t^2). \tag{6.10}$$

Let us review how this expression comes about. The piece $1 + t^2 + 4t^3 + t^4 + t^6$ is the contribution of forms supported at $\mu^{-1}(0)$, and the piece $t^4(1 + 34t + t^2)$ is the contribution of forms supported at $\mu^{-1}(\frac{\pi}{2})$ (using the notation from §6.5.2). Thus, the polynomial $1 + t^2 + 4t^3 + t^4 + t^6$ is the Poincaré polynomial of the moduli space $\mathcal{M}_{\text{fc}}$ of flat $SU(2)$ connections over a genus $h = 2$ Riemann surface with one unit of magnetic flux.

The cohomology of $\mathcal{M}_{\text{fc}}$ for $SU(2)$ with one unit of magnetic flux and genus $h > 1$ has been calculated in [52] (see also [53]). It has $2h + 2$ generators: $\alpha \in H^2(\mathcal{M}_{\text{fc}})$, $\beta_1, \ldots, \beta_{2h} \in H^3(\mathcal{M}_{\text{fc}})$, and $\gamma \in H^4(\mathcal{M}_{\text{fc}})$. Let us briefly review where these generators come from. From a flat connection over $\mathcal{C}_h$ one can construct a holomorphic rank-2

vector bundle over $\mathcal{C}_h$. Since this vector bundle varies as a function of the point in $\mathcal{M}_{\text{fc}}$ we get a vector bundle over $\mathcal{C}_h \times \mathcal{M}_{\text{fc}}$. The second Chern class c_2, which is in $H^4(\mathcal{C}_h \times \mathcal{M}_{\text{fc}})$ can be decomposed in terms of a basis of $H^*(\mathcal{C}_h)$. The coefficient of the generator of $H^0(\mathcal{C}_h)$ is γ, the coefficient of the generator of $H^2(\mathcal{C}_h)$ is α, and the coefficients of the $2h$ generators of $H^1(\mathcal{C}_h)$ are the β_j's. There is a quite complicated set of relations [54] among $\alpha, \beta_1, \ldots, \beta_{2h}, \gamma$, which we will not need in the present paper. The case of genus $h = 2$ is particularly easy to describe. In this case, by Hodge duality we can complete the Poincaré polynomial of $\mathcal{M}_{\text{fc}}$ to $1 + t^2 + 4t^3 + t^4 + t^6$.

Another useful fact is that the mapping class group of $\mathcal{C}_h$ acts nontrivially on the generators $\beta_1, \ldots, \beta_{2h}$. The mapping class group is the fundamental group of the space of complex structures of $\mathcal{C}_h$. As one traverses a loop in this space, the complex structure of $\mathcal{C}_h$ varies and with it the space $\mathcal{M}_{\text{fc}}$ varies. As one completes the loop, the complex structure of $\mathcal{C}_h$ is back to its original value, and the space $\mathcal{M}_{\text{fc}}$ is also isomorphic to the original space at the start of the loop, but a particular generator of $H^3(\mathcal{M}_{\text{fc}})$ does not necessarily map to itself. In general, there is a nontrivial action [described by an element in the symplectic group $\mathrm{Sp}(2h, \mathbb{Z})$ which preserves the intersection form] on the generators of $H^1(\mathcal{C}_h)$, which induces a nontrivial dual action on $\beta_1, \ldots, \beta_{2h}$.

The remainder of the Poincaré polynomial (6.10) is the contribution from reducible solutions of the form (6.7) with $\alpha_z \neq 0$. The cohomology that we need is the subspace invariant under large gauge transformations, as discussed at the end of §6.5.2, and the corresponding Poincaré polynomial is

$$P(t) = 1 + t^2 + 4t^3 + t^4 + t^6 + t^4(1 + 4t + t^2). \tag{6.11}$$

The space $\mathcal{M}_H$ is noncompact, so we need to specify whether we allow forms with noncompact support. Since these correspond to nonnormalizable states, we will drop them, and so we work with the cohomology with compact support. Let us denote by $\delta(\mu^{-1}(0))$ the 6-form with support on $\mu^{-1}(0)$ and "indices" in the direction transverse to $\mu^{-1}(0)$ which are the directions of the base of the Hitchin fibration. [$\delta(\mu^{-1}(0))$ can be smeared out to what is known as the *Thom class* of a tubular neighborhood of $\mu^{-1}(0)$.] Let us also denote by $\delta(\mu^{-1}(\frac{\pi}{2}))$ the 10-form with support on $\mu^{-1}(\frac{\pi}{2})$ (which is a Riemann surface and therefore has 10 orthogonal directions). For the cohomology with compact support we have to multiply the piece $1 + t^2 + 4t^3 + t^4 + t^6$ in (6.11) by t^6. We get representatives of the cohomology on $\mathcal{M}_H$ by multiplying the corresponding forms on $\mathcal{M}_{\text{fc}}$ by $\delta(\mu^{-1}(0))$. Similarly, $\mu^{-1}(\frac{\pi}{2})$ is a Riemann surface and has a Poincaré polynomial $1 + 4t + t^2$ (ignoring the complication of the 16-fold cover mentioned at the end of §6.5.2). We get the corresponding forms on $\mathcal{M}_H$ by multiplying the forms on $\mu^{-1}(\frac{\pi}{2})$ by $\delta(\mu^{-1}(\frac{\pi}{2}))$.

6.5.4 Action of γ

The R-symmetry twist γ acts on ϕ_z as

$$\phi_z \to e^{iv}\phi_z,$$

according to (2.11). It therefore acts on the quadratic differential as

$$b_{zz} \to e^{2iv} b_{zz}\,.$$

Note that $e^{2iv} \neq 1$ for all the values of v listed in §2.1. It follows that a γ-invariant point of $\mathcal{M}_H$ is possible only if $b_{zz} = 0$. Thus, the only fixed points of γ occur over the singular fiber of the Hitchin fibration. This conclusion holds for any of the values of n and v from §2.3.

Restricting to the singular fiber over $b_{zz} = 0$, the γ-invariant subspace is the disjoint union $\mu^{-1}(0) \cup \mu^{-1}(\frac{\pi}{2})$, i.e., the union of the moduli space of flat connections $\mathcal{M}_{\rm fc}$ and solutions with an abelian gauge field ($c_{\overline{z}} = 0$), which is a copy of $\mathcal{C}_2$ (see the notation at the end of §6.5.2). To see this, note that $\phi_z = 0$ is obviously γ-invariant. This gives $\mu^{-1}(0)$. For nonzero ϕ_z of the form (6.7) we have

$$e^{iv}\phi_z = \begin{pmatrix} e^{\frac{1}{2}iv} & 0 \\ 0 & e^{-\frac{1}{2}iv} \end{pmatrix} \phi_z \begin{pmatrix} e^{\frac{1}{2}iv} & 0 \\ 0 & e^{-\frac{1}{2}iv} \end{pmatrix}^{-1}.$$

This gauge transformation, however, doesn't preserve $c_{\overline{z}}$, and only $c_{\overline{z}} = 0$ solutions, i.e., those in $\mu^{-1}(\frac{\pi}{2})$ are γ-invariant.

6.5.5 S-duality

As shown in [6], the coupling constant of the four-dimensional gauge theory determines the Kähler structure of $\mathcal{M}_H$ upon compactification on $\mathcal{C}_h$, which is also the Kähler structure of each fiber. S-duality in four dimensions therefore becomes the fiberwise T-duality of the two-dimensional σ-model.

We need to understand the action of $\mathrm{SL}(2,\mathbb{Z})$-duality on the A-model operators, i.e., on the de Rham cohomology of $\mathcal{M}_H$. In this subsection we may assume that the σ-model is formulated on $\mathbb{R}^2$. In principle, the action of S-duality on a generic fiber of the Hitchin fibration is tractable, as it reduces to T-duality on the T^{6h-6} fiber. However, it is not so clear how to track this action to the singular fiber, which is what we need. We will thus employ a few indirect arguments. We will also restrict ourselves to the case $h = 2$ and only the S-duality element $\tau \to -1/\tau$. We will thus keep denoting it by $\mathcal{S}$ but restrict to $v = \pi/2$.

In this section it will be more convenient to work with the cohomology $H^*(\mathcal{M}_H)$ with noncompact support, rather than the cohomology of forms with compact support

$H^*_{\rm cpt}(\mathcal{M}_H)$. Since the fibers of the Hitchin fibration are compact, the distinction between compact and noncompact cohomology only depends on the behavior of the forms as a function of the base point. S-duality preserves the base point, and in a sense acts classically on the base. We lose no information by working with noncompact forms.

We can then use the following facts:

(i) $\mathcal{S}^2$ acts as charge conjugation on the gauge theory. For $SU(n)$ gauge group charge conjugation acts on the gauge field as $A \to -A^t$ and on the Higgs field as $\phi_z \to \phi_z^t$, where $(\cdots)^t$ is the transpose operation. Combining $\mathcal{S}$ with the R-symmetry twist γ from (2.11) (with $\upsilon = \pi/2$) we find that $(\mathcal{S}\gamma)^2$ acts as:

$$(\mathcal{S}\gamma)^2 : \qquad A \to -A^t\,, \qquad \phi_z \to -\phi_z^t.$$

For $\mathfrak{su}(n)$ with $n > 2$ the automorphism $x \to -x^t$ is outer, but for $\mathfrak{su}(2)$ it is an inner automorphism, as $-x^t = \sigma_2^{-1} x \sigma_2$ where $\sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \in SU(2)$. Thus, for an $SU(2)$ gauge group, $(\mathcal{S}\gamma)^2$ is equivalent to a gauge transformation and acts as the identity on the A-model operators. It follows that the eigenvalues of $\mathcal{S}\gamma$ are ± 1.

(ii) Both $\mathcal{S}$ and γ commute with the mapping class group $\mathrm{Sp}(2h,\mathbb{Z})$ of $\mathcal{C}_h$. The mapping class group $\mathrm{Sp}(2h,\mathbb{Z})$ acts nontrivially on the cohomology of $\mathcal{M}_H$. The action of $\mathrm{Sp}(2h,\mathbb{Z})$ is generated by operations that can be described as follows. Suppose we cut $\mathcal{C}_h$ along a one-cycle (which we identify with S^1) and glue it back with a rotation by an angle θ (a Dehn twist), understood as part of a holomorphic transformation in a local neighborhood of the cut. This defines a new complex structure on $\mathcal{C}_h$, and as we let θ vary continuously from 0 to 2π we get a loop in the moduli space of complex structure of $\mathcal{C}_h$. As we traverse the loop, we can follow what happens to an integral cohomology class of $\mathcal{M}_H$, and after the loop is completed we generally find that it is not back to itself. In this way we get a nontrivial action on $H^*(\mathcal{M}_H)$. We can therefore decompose $H^*(\mathcal{M}_H)$ into irreducible representations of $\mathrm{Sp}(2h,\mathbb{Z})$, and $\mathcal{S}\gamma$ has to act as either the identity or multiplication by (-1) in each irreducible subspace. (If an irreducible representation of $\mathrm{Sp}(2h,\mathbb{Z})$ appears in the decomposition of $H^*(\mathcal{M}_H)$ with multiplicity higher then 1, then $\mathcal{S}\gamma$ can mix these subspaces, but we can always diagonalize it in the direct sum of these subspaces and the eigenvalues will be ± 1.)

(iii) The A-model is independent of the complex structure of the target space altogether, and only the Kähler class is important. Thus, $\mathcal{S}\gamma$ is invariant under complex conjugation.

(iv) Some of the operators of the A-model are directly related to the topological operator $\mathcal{O} = \int F \wedge F$ of $\mathcal{N} = 4$ SYM. At the selfdual point $\tau = i$ it is not hard to check that S-duality acts as $\mathcal{O} \to -\mathcal{O}$ (a small θ-angle is mapped to its negative). In order to understand to which operators of the σ-model this observation is relevant, recall [37] that a local operator $\mathcal{O}^{(0)}$ of the A-model has a nonlocal descendant $\mathcal{O}^{(2)}$ that can be expressed as an integral over all space $\mathbb{R}^2$. Let $\mathcal{O}^{(0)}$ be the operator associated with the cohomology class in $H^2(\mathcal{M}_H)$ that descends from $\alpha \in H^2(\mathcal{M}_{\mathrm{fc}})$ (discussed in §6.5.3). (Note that $\mathcal{O}^{(0)}$ does not correspond to a compactly supported class, since we haven't multiplied it yet by the 6-form $\delta(b_{zz})$, but this is unnecessary for the purposes of understanding the action of S-duality.) Since α was defined in terms of the second Chern class c_2, by definition $\mathcal{O}^{(2)}$ is proportional to $\int F \wedge F$ on the entire space. We conclude that $\mathcal{S}\gamma$ acts as (-1) on α. Applying a similar argument to the β_j's in $H^3(\mathcal{M}_{\mathrm{fc}})$ and $\gamma \in H^4(\mathcal{M}_{\mathrm{fc}})$ runs into minor difficulties, since the physical $F \wedge F$ can only be reduced to a 2-form on $\mathbb{R}^2$ (by integration on $\mathcal{C}_h$), and, geometrically, the σ-model induced map can only turn β_j and γ into 3-forms and 4-forms respectively. But the A-model descendent $\mathcal{O}^{(2)}$ that corresponds to, say, γ is an integral of a 2-form on $\mathbb{R}^2$ and contains two fermionic fields of the A-model.

(v) Once we have $\mathcal{S}\gamma(\alpha) = -\alpha$ we can use the cohomology product to obtain $\mathcal{S}\gamma(\alpha^2) = \alpha^2$ and $\mathcal{S}\gamma(\alpha^3) = -\alpha^3$. Here it is crucial to work in cohomology with noncompact support, since the product is known to be trivial for the cohomology with compact support $H^*_{\mathrm{cpt}}(\mathcal{M}_H)$ [55]. As for the cohomology with noncompact support, α is a 2-form, which must therefore be proportional to the Kähler class of $\mathcal{M}_H$ (in complex structure I), since $H^2(\mathcal{M}_H)$ is 1-dimensional. Therefore α^2 and α^3 are a nonzero 4-form and 6-form, respectively.

(vi) We can gather extra clues from the assumption that S-duality acts as a simple T-duality on the generic T^6 fiber of the Hitchin fibration [6]. We have seen in §4.4 that T-duality acts as multiplication by i^{p+q} on the operators of the A-model that correspond to elements in the $H^{(p,q)}$ of Dolbeault cohomology. This was shown for T^2, but the result clearly generalizes to T^6. This operation does not square to the identity, but recall from §6.5.4 that the R-symmetry twist γ acts nontrivially on the base of the Hitchin fibration: $b_{zz} \to -b_{zz}$. Thus, when discussing the action of $\mathcal{S}\gamma$ on a generic fiber, we have to consider both the fiber at b_{zz} and the fiber at $-b_{zz}$ simultaneously. Let F be the fiber over b_{zz} and F' be the fiber over $-b_{zz}$. Since b_{zz} and $-b_{zz}$ have the same zeroes over $\mathcal{C}_2$, it follows (by definition of the fibers as the moduli space of flat connections over a Riemann surface that is the

double cover of $\mathcal{C}_2$ branched over the zeroes of b_{zz}) that F and F' are naturally isomorphic. $\mathcal{S}\gamma$ interchanges F and F', and takes the block-form:

$$\begin{pmatrix} 0 & \imath^{p+q} \\ (-i)^{p+q} & 0 \end{pmatrix}, \tag{6.12}$$

where the first block of columns or rows refers to the $H^{(p,q)}(F)$ and the second block refers to $H^{(p,q)}(F')$. We have replaced $\imath^{p+q}$ with $(-i)^{p+q}$ in the second block so as to keep $(\mathcal{S}\gamma)^2 = 1$. What can we learn from this about the action of $\mathcal{S}\gamma$ on $\mathcal{M}_H$? The inclusion maps $\imath : F \hookrightarrow \mathcal{M}_H$ and $\imath' : F' \hookrightarrow \mathcal{M}_H$ induce maps on cohomology $\imath^* : H^*(\mathcal{M}_H) \to H^*(F)$ and $\imath'^* : H^*(\mathcal{M}_H) \to H^*(F')$. And $\mathcal{S}\gamma$ commutes with these maps, in the sense that $\imath'^* \circ (\mathcal{S}\gamma) = (\mathcal{S}\gamma) \circ \imath^*$. We can identify $H^*(F) \simeq H^*(F')$ and write $\imath^* \circ (\mathcal{S}\gamma) = (\mathcal{S}\gamma) \circ \imath^*$. Thus, if $\lambda \in H^{(p,q)}(\mathcal{M}_H)$ we get $\imath^* \mathcal{S}\gamma(\lambda) = (-i)^{p+q} \imath^*(\lambda)$. This doesn't uniquely determine $\mathcal{S}\gamma(\lambda)$ since $\imath^*$ might not be injective, but it gives us partial information. For example, $\imath^*$ is injective on $H^{(1,1)}(\mathcal{M}_H)$ since it maps the Kähler class of $\mathcal{M}_H$ (in complex structure I) to the Kähler class of F. So, we again recover the result that $\mathcal{S}\gamma$ acts as $(-i)^{p+q} = -1$ on the 2-form α.

To summarize, at this point we have the following information for $h = 2$:

- On the 1-dimensional $H^0(\mathcal{M}_H)$ (or $H^6_{\text{cpt}}(\mathcal{M}_H)$) $\mathcal{S}\gamma$ acts as $+1$;
- On the 1-dimensional $H^2(\mathcal{M}_H)$ (or $H^8_{\text{cpt}}(\mathcal{M}_H)$) $\mathcal{S}\gamma$ acts as -1;
- On the 4-dimensional $H^3(\mathcal{M}_H)$ (or $H^9_{\text{cpt}}(\mathcal{M}_H)$) $\mathcal{S}\gamma$ has either 4 eigenvalues of -1 or 4 eigenvalues of $+1$.
- On the 2-dimensional $H^4(\mathcal{M}_H)$ (or $H^{10}_{\text{cpt}}(\mathcal{M}_H)$) $\mathcal{S}\gamma$ has one eigenvalue $+1$ and the other eigenvalue is either $+1$ or -1.
- On the 4-dimensional $H^5(\mathcal{M}_H)$ (or $H^{11}_{\text{cpt}}(\mathcal{M}_H)$) $\mathcal{S}\gamma$ has either 4 eigenvalues of -1 or 4 eigenvalues of $+1$.
- On the 2-dimensional $H^6(\mathcal{M}_H)$ (or $H^{12}_{\text{cpt}}(\mathcal{M}_H)$) $\mathcal{S}\gamma$ has one eigenvalue -1 and the other eigenvalue is either $+1$ or -1.

Thus, we know the action of $\mathcal{S}\gamma$ up to four undetermined ($\pm$) signs.

6.6 Effect of fluxes: sharpening the conjecture

We still need to explain which gauge bundle to take for our conjectured low-energy Chern–Simons theory, i.e., what is the three-dimensional magnetic flux. To answer this question we will now consider the effect of electric and magnetic fluxes of the four-dimensional theory on the two-dimensional σ-model. This will lead us to a sharpened version of our conjecture. The reader who is not interested in the details, but nonetheless trusts the authors, is advised at this point to skip to the last sentence of this subsection.

Let us first consider the case with $\tau = i$ and $\mathbf{s} = \mathbf{s}'$. We saw in §3.5 that in this case, among the Hilbert spaces $\mathcal{H}_{\mathbf{e},\mathbf{m}}$—each associated with a choice of electric and magnetic fluxes $\mathbf{e}$ and $\mathbf{m}$—the only ones that are invariant under the S-duality action of $\mathbf{s}'$ are those with $\mathbf{e} = \mathbf{m}$. We will therefore focus our attention on these subspaces.

Since the four-dimensional theory is defined on the space $X = S^1_R \times \mathcal{C}_h \times S^1_T$, where the first S_1 is regarded as Euclidean time, we can decompose the fluxes further in the following way [21]:

$$\mathbf{e} = \mathbf{e}_0 + \mathbf{e}_1 \in \mathbb{Z}_2 \oplus H^1(\mathcal{C}_h, \mathbb{Z})\,, \tag{6.13}$$

$$\mathbf{m} = \mathbf{m}_0 + \mathbf{m}_1 \in \mathbb{Z}_2 \oplus H^1(\mathcal{C}_h, \mathbb{Z})\,. \tag{6.14}$$

Roughly speaking, $\mathbf{e}_0$ and $\mathbf{e}_1$ are electric fluxes along S^1_T and a one-cycle of $\mathcal{C}_h$, respectively, and $\mathbf{m}_0$ and $\mathbf{m}_1$ are magnetic fluxes through $\mathcal{C}_h$ and a two-cycle consisting of S^1_R and a one-cycle of $\mathcal{C}_h$.

As we reviewed in the previous section, we need nonzero $\mathbf{m}_0$ in order to have a smooth moduli space $\mathcal{M}_H$. Therefore, we also choose nonzero $\mathbf{e}_0$ to have an $\mathbf{s}'$-invariant Hilbert space. After compactifying on $\mathcal{C}_h$, nonzero $\mathbf{e}_0$ implies, according to [21], the presence of a flat B-field in the sigma model.

This leaves us the freedom of choice of $\mathbf{e}_1 = \mathbf{m}_1$. To interpret these in the two-dimensional terms, we need to distinguish between the moduli spaces $\mathcal{M}_H(\mathcal{C}_h, G)$ for different gauge groups $G = SU(2)$ and $G = SO(3)$. The moduli space for $SU(2)$ was briefly described in §6.5. It is also shown in [50] that the space is simply connected. On the other hand, it possesses a geometric symmetry group $H^1(\mathcal{C}_h, \mathbb{Z}_2)$ (which acts by changing the holonomies around the one-cycles by elements of $\mathbb{Z}_2$), and upon dividing the space by this symmetry, we get the moduli space for $SO(3)$, whose fundamental group is $H^1(\mathcal{C}_h, \mathbb{Z}_2)$.

It is now clear that states of the σ-model with target space $\mathcal{M}_H(SU(2))$ will carry conserved momenta corresponding to the symmetry group $H^1(\mathcal{C}_h, \mathbb{Z}_2)$, while strings of the sigma model with target space $\mathcal{M}_H(SO(3))$ will carry winding numbers valued in its fundamental group $H^1(\mathcal{C}_h, \mathbb{Z}_2)$. The quantities $\mathbf{e}_1$ and $\mathbf{m}_1$ signify the conserved

momentum and winding number of the respective σ-models, and the fact that they are exchanged under the T-duality is another manifestation of the fact that S-duality of four-dimensional gauge theory reduces to T-duality of the two-dimensional σ-model upon compactification [6, 21]. Furthermore, it is shown in [56] that the two moduli spaces $\mathcal{M}_H(SU(2))$ and $\mathcal{M}_H(SO(3))$ are indeed mirror pairs.

It is equally clear, however, that we cannot have nonzero $\mathbf{e}_1$ and $\mathbf{m}_1$ at the same time in the two-dimensional sigma model; they simply correspond to strings living in two different target spaces. The only way to achieve $\mathbf{e}_1 = \mathbf{m}_1$ is to set them both to zero. States with $\mathbf{e}_1 = 0$ in the σ-model with target space $\mathcal{M}_H(SU(2))$ are invariant under the action of the symmetry group $H^1(\mathcal{C}_h, \mathbb{Z}_2)$, so they descend to well-defined states after dividing the moduli space by the symmetry group to make the target space $\mathcal{M}_H(SO(3))$. On the other hand, strings moving in $\mathcal{M}_H(SO(3))$ with zero winding numbers can be lifted to strings moving in the covering space $\mathcal{M}_H(SU(2))$. [6] Therefore, the condition $\mathbf{e}_1 = \mathbf{m}_1 = 0$ is consistent.

These considerations lead us to the following formulation of our conjecture. We start with a four-dimensional gauge theory on $X = S^1_R \times \mathcal{C}_h \times S^1_T$, with S-duality and R-symmetry twists inserted at a point of S^1_R. If we focus on the sector with $\mathbf{e}_0 = \mathbf{m}_0 = 1$ and $\mathbf{e}_1 = \mathbf{m}_1 = 0$, then in the limit of small $\mathcal{C}_h$ (limit (ii) of §6.4), where the theory becomes a σ-model with target space $\mathcal{M}_H(SO(3))$, we end up with the sector with zero winding number of the σ-model. S-duality becomes T-duality of the σ-model, which sends the states with $\mathbf{m}_1 = 0$ to states with $\mathbf{e}_1 = 0$ of the σ-model whose target space is now the universal cover $\mathcal{M}_H(SU(2))$. But states with $\mathbf{e}_1 = 0$ are precisely those that can be interpreted as states of the σ-model with target space $\mathcal{M}_H(SO(3))$, as explained in the previous paragraph, and hence the action of T-duality is well-defined in this sector.

On the other hand, in the limit where S^1_R shrinks to a point (limit (i) of §6.4), we conjecture that the theory becomes a Chern–Simons theory on $M_3 = \mathcal{C}_h \times S^1_T$. Therefore, the partition function of Chern–Simons theory will calculate the trace of the T-duality times an R-symmetry operator on the Hilbert space of the σ-model. Now, the $SO(3)$-bundles on M_3 are classified according to their "magnetic fluxes" $\mathbf{m} \in H^2(M_3, \mathbb{Z}_2)$, which decomposes as $\mathbf{m} = \mathbf{m}_0 + \mathbf{m}_1$ in exactly same way as (6.14). Hence the partition

[6] In fact, a string in $\mathcal{M}_H(SO(3))$ can be lifted to many copies of strings in $\mathcal{M}_H(SU(2))$, related to each other by the action of the symmetry group. The precise statement here is that a state of zero winding number of the σ-model with target space $\mathcal{M}_H(SO(3))$ will lift to a unique state of the σ-model with target space $\mathcal{M}_H(SU(2))$ with zero momentum (i.e., invariant under the action of the symmetry group).

function can be written as

$$Z = \sum_{\mathbf{m}_0, \mathbf{m}_1} Z_{\mathbf{m}_0, \mathbf{m}_1} \, .$$

Comparing the two descriptions in different limits, it is natural to conclude that

$$I_{1,0} = Z_{1,0} \, ,$$

where $I_{1,0}$ is the index, defined in (6.5), of the σ-model with target space $\mathcal{M}_H(SO(3))$ restricted to the zero winding number sector, and $Z_{1,0}$ is the contribution to the Chern–Simons theory partition function from bundles with $\mathbf{m}_0 = 1$ and $\mathbf{m}_1 = 0$. This is the sharpened version of our conjecture.

For the other two cases where $\tau = e^{\pi i/3}$ and $\mathbf{s} = \pm \mathbf{s}''$, we saw in §3.5 that $\mathbf{s}$-invariance of the Hilbert space $\mathcal{H}_{\mathbf{e},\mathbf{m}}$ forces us to choose $\mathbf{e} = \mathbf{m} = 0$. This means that the target space of the σ-model that we get upon compactification on $\mathcal{C}_h$ is a singular one. We could avoid dealing with singular target space by inserting Wilson/'t Hooft operators as discussed at the end of §3.5. In the opposite limit, we would then have to calculate the expectation values of these line operators in the Chern–Simons theory, instead of its partition function. We leave these possibilities for a future work, and concentrate on $\tau = i$ and $\mathbf{s} = \mathbf{s}'$ case for now.

6.7 Testing the conjecture

We are now ready to compare the two limits of counting vacua. In limit (i) $\mathcal{C}_h$ is large and we first reduce on S^1_R. By the conjecture of §6.1 the result is Chern–Simons theory at level k. There are no low-energy fields left that carry R-charge, so the twist γ has no effect. The partition function is simply $d_h(n, k, \mathbf{m}_0)$ — the dimension of the Hilbert space of $SU(n)$ Chern–Simons theory at level k on $\mathcal{C}_h$ with the magnetic flux specified by $\mathbf{m}_0$. For $SU(2)$ with one unit of magnetic flux $\mathbf{m}_0 = 1$, genus $h = 2$, and level $k = 2$ (corresponding to S-duality $\tau \to -1/\tau$ and $\tau = i$) we get $d_2(2, 2, 1) = 6$ (see Appendix A for details).

In limit (ii), $\mathcal{C}_h$ is small and we first reduce on it to obtain a supersymmetric σ-model with Hitchin's space $\mathcal{M}_H$ as the target space. We then compactify that on S^1_R with a T-duality twist and an R-symmetry twist γ. The latter acts only on the Higgs fields $\phi_z, \overline{\phi}_{\overline{z}}$ in (6.6). We need the (absolute value of the) supertrace of $\mathcal{S}\gamma$ on the Hilbert space of the A-model compactified on S^1 (i.e., S^1_T). For this, we need to know the action of $\mathcal{S}\gamma$ on the A-model states, which are in one-to-one correspondence with the cohomology of $\mathcal{M}_H$ with compact support. As reviewed in §6.5, $\mathcal{M}_H$ has a fibration structure with the base B being the moduli space of gauge-invariant polynomials in ϕ_z. The twist γ acts on that space, and by our restrictions on the rank n and the

discussion in §2.3 we may assume that the only fixed point of γ in B is the origin of the Hitchin fibration where $b_{zz} = \mathrm{tr}(\phi_z^2) = 0$ (see the discussion above in §6.5.4 for more details). This leaves the singular fiber, which is compact, and so the partition function is well defined. This is the reason why we restrict to the cohomology with compact support. In fact, we may just as well restrict to elements of cohomology that are supported on the singular fiber. Now we can collect the information from §6.5.5 and attempt to reproduce in limit (ii) the number 6 that we got in limit (i). We have to calculate the alternating sum of traces of $\mathcal{S}\gamma$ in the subspaces $H^i(\mathcal{M}_H)$. Unfortunately, there are several signs that we did not determine in §6.5.5, and so our conjecture that the sum is 6 cannot be tested at this point. However, it is easy to see that there are several ways to choose the undetermined signs so as to reproduce the required result 6, so at this point our conjecture cannot be ruled out either. In principle, given the exact expressions for the representatives of the cohomology of $\mathcal{M}_H$, it is possible to calculate the action of $\mathcal{S}$ on them using the general framework of [57, 58, 59], but this is beyond the scope of this paper.

6.8 A six-dimensional perspective

We will now briefly comment on some aspects of our construction that can be understood better in terms of the (2,0)-theory. We mentioned the six-dimensional realization of our setting, in terms of the (2,0)-theory compactified on T^2, in (2.4)-(2.5). The identification (2.4) takes care of the $\mathrm{SL}(2,\mathbb{Z})$-twist, but the R-symmetry twist needs to be added as well. The R-symmetry group of the (2,0)-theory is Sp(4) [the double cover of $SO(5)$]. While the $\mathcal{N}=6$ R-symmetry twist (2.11) cannot be embedded in Sp(4), the $\mathcal{N}=4$ twist (2.12) can be, if $e^{i\varphi_4} = \pm e^{-\frac{1}{2}i\upsilon}$.

The lift to six-dimensions introduces a new dimensionful parameter — the area $\mathcal{A}$ of T^2, which has to be taken to zero before all other limits (small R or small $\mathcal{C}_h$) are taken. So far we considered two different limits: one in which the size of $\mathcal{C}_h$ is large compared to the size R of S^1, and the other in which S^1 is large compared to $\mathcal{C}_h$. We now find yet another possibly interesting limit to consider. In this limit we take the scale $\sqrt{\mathcal{A}}$ of T^2 to be much larger than R. In the limit $R \ll \sqrt{\mathcal{A}}$ it is more useful to think about the space given by (2.4) as a circle fibration over an orbifold of $T^2/\mathbb{Z}_q$ given by $z \sim e^{i\upsilon}z$ ($q = 4, 6$, or 3, according to whether $\upsilon = \frac{\pi}{2}, \frac{\pi}{3}$, or $\upsilon = \frac{4\pi}{3}$). The radius of the S^1 fiber is Rq and the structure group is $\mathbb{Z}_q$. The base $T^2/\mathbb{Z}_q$ has several orbifold points, which are solutions of $z = e^{i\upsilon}z + n + m\tau$ for some $n, m \in \mathbb{Z}$:

- For $q = 4$ (and $\tau = i$) there are 3 fixed points: $z = 0$ and $z = (1+i)/2$ both with monodromy $\mathbb{Z}_4$, and $z = \frac{1}{2}$ with monodromy $\mathbb{Z}_2$ (since $z = \frac{1}{2}$ is not fixed by the rotation $z \to e^{i\upsilon}z$ but is fixed by $z \to e^{2i\upsilon}z$).

- For $q = 6$ (and $\tau = e^{\frac{\pi i}{3}}$) there are 3 inequivalent fixed points: $z = 0$ with monodromy $\mathbb{Z}_6$, $z = \frac{1}{2} + \frac{i}{2\sqrt{3}}$ with monodromy $\mathbb{Z}_3$, and $z = \frac{1}{2}$ with monodromy $\mathbb{Z}_2$.

- For $q = 3$ (and $\tau = e^{\frac{\pi i}{3}}$) there are 3 inequivalent fixed points: $z = 0$, $z = \frac{1}{2} + \frac{i}{2\sqrt{3}}$ and $z = \frac{i}{\sqrt{3}}$, all with monodromy $\mathbb{Z}_3$.

The effective description as $R \to 0$ is weakly coupled 4+1D $\mathcal{N} = 2$ SYM, with coupling constant $g_{\text{YM}}^2 = 8\pi^2 Rq$, on $\mathbb{R}^{2,1} \times (T^2/\mathbb{Z}_q)$. The only complication arises from the fixed points of $\mathbb{Z}_q$ listed above. To understand the behavior of the theory near each fixed point we can replace T^2 with $\mathbb{R}^2$. We are thus led to study the following question: what is the effective low-energy description of the $(2,0)$-theory formulated on $\mathbb{R}^{2,1} \times [(\mathbb{C} \times S^1)/\mathbb{Z}_q]$ where $[(\mathbb{C} \times S^1)/\mathbb{Z}_q]$ is the orbifold of $\mathbb{C} \times S^1$ (parameterized by (z, x_3) with $0 \leq x_3 < 2\pi Rq$) by $\mathbb{Z}_q$ that is generated by (the freely acting) $(z, x_3) \mapsto (e^{2\pi i q} z, x_3 + 2\pi R)$? We can also add an R-symmetry transformation γ of order q to the above action.

The requisite low-energy description should be formulated on $\mathbb{R}^{2,1} \times (\mathbb{C}/\mathbb{Z}_q)$ and should be 4+1D $\mathcal{N} = 2$ SYM away from the origin of $\mathbb{C}$. The question is what are the extra (2+1D) modes that are localized at the origin. This setting is reminiscent of the Melvin background, and D-branes in this background have been studied extensively [60]. Furthermore, using the M-theory realization of the $(2,0)$-theory as the low-energy theory of M5-branes, and the relation between the geometry discussed above and the M-theory lift of (p,q) 5-branes [61], one can relate the three-dimensional boundary degrees of freedom at the singular point of $\mathbb{C}/\mathbb{Z}_q$ to the boundary degrees of freedom at the intersection of D3-branes with $(1,q)$ 5-branes, which were recently solved in [62][13]. Part of the answer is Chern–Simons theory at the fractional level $k = 1/q$ [63]. This can be argued by noting that the bulk action contains a term of the form $\frac{1}{2\pi q} \int F \wedge F$ where the integral is over $\mathbb{R}^{2,1} \times C$ and $C \subset \mathbb{C}/\mathbb{Z}_q$ is any open path from $z = 0$ to $z = \infty$. This term can be integrated to give a Chern–Simons term at level $k = \frac{1}{q}$ at the origin (minus a similar term at infinity). Thus Chern–Simons actions naturally arise at this limit $\sqrt{\mathcal{A}} \gg R$ as well, although we have to note that the gauge field variables in this limit do not have a direct relation to the gauge field variables in the opposite limit $R \gg \sqrt{\mathcal{A}}$.

7. Discussion

We have put forward various arguments that suggest that a three-dimensional topological structure underlies S-duality of $\mathcal{N} = 4$ SYM. By "structure" we mean a (probably

nonlocal) action $\mathcal{S}(V, \widetilde{V})$ that depends on two independent gauge field configurations and their supersymmetric partners, but is independent of the metric or coupling constant. Here V is the "original" gauge field configuration A together with the superpartners, and $\widetilde{V}$ is the "dual" gauge field configuration $\widetilde{A}$ together with its superpartners. In other words, we used the Fredholm kernel representation of an operator in four dimensions, to construct the action of a field theory in three dimensions. It is interesting to wonder whether this setting is related to a more general framework put forward recently in [64] whereby a wavefunction of a quantum field theory in d dimensions is related to the action functional of another quantum field theory in one dimension less.[7] We have also argued that, with appropriate modifications and restrictions, $\mathcal{S}(V, \widetilde{V})$ leads to a *local* topological field theory when we set $V = \widetilde{V}$. This field theory had a direct, local description in terms of a twisted circle compactification. We conjectured that it is a Chern–Simons theory.

In §6.4 we studied the Witten index of the supersymmetric compactification on a Riemann surface $\mathcal{C}_h$ times an S^1 with an R-twist and an $\mathrm{SL}(2, \mathbb{Z})$-twist. We considered two different limits: one in which the size of $\mathcal{C}_h$ is large compared to the size of S^1, and the other in which S^1 is large compared to $\mathcal{C}_h$. When $\mathcal{C}_h$ is large, we could use our conjecture about the relation to Chern–Simons theory to calculate the number of vacua. When $\mathcal{C}_h$ is small, we used the topological string theory on Hitchin's space to calculate the Witten index in the special case of $SU(2)$ gauge group with an appropriate flux, $\tau \to -1/\tau$ twist at $\tau = i$, and genus $h = 2$. But we fell short of a full comparison, because we did not determine several $(\pm)$ signs in the action of S-duality. As we have argued in §6.7, some sign assignments are consistent with our conjecture, and some are not. It would obviously be interesting to establish these signs and also extend the tests to higher genus, other gauge groups, and other $\mathrm{SL}(2, \mathbb{Z})$ elements.

It would also be interesting to understand in more detail the T-duality (mirror-symmetry) twist and its relation to geometric quantization, as discussed in §4.4. The general question here is what is the low-energy description of a σ-model that is selfdual under mirror symmetry when compactified on S^1 with a mirror symmetry twist. The simple examples in §4.2 suggest that the answer is related to geometric quantization of the target space. A more general problem can involve a twist by a combination of mirror symmetry and a geometrical isometry γ. It is then interesting to explore the relation between the 0+1D low-energy description (i.e., the low-energy Hilbert space) and geometric quantization of the γ-invariant subspace of target space. In the context of our setting of §6, the selfdual target space is Hitchin's space $\mathcal{M}_H$ associated with $\mathcal{C}_h$, and the low-energy 0+1D theory is Chern–Simons theory compactified on $\mathcal{C}_h$,

[7]We wish to thank Petr Hořava for suggesting this connection.

which can be identified with geometric quantization of $\mathcal{M}_{\rm fc}$ — the moduli space of flat connections on $\mathcal{C}_h$. However, $\mathcal{M}_{\rm fc}$ is not quite the γ-invariant subspace of $\mathcal{M}_H$. In the case we considered of gauge group $SU(2)$, the γ-invariant subspace is actually a disjoint union of $\mathcal{M}_{\rm fc}$ and a copy of $\mathcal{C}_2$ (see §6.5.4 above).

Returning to the general case of a selfdual σ-model, the T-duality (mirror symmetry) twist treats differently the left and right moving modes of the σ-model. One would therefore like to analyze separately the left and right moving CFTs with the twist. One tool that might prove useful in this analysis is the recent construction of Frenkel, Losev and Nekrasov [65], where the complex structure τ of $\mathcal{N}=4$ SYM is treated independently from its complex conjugate $\overline{\tau}$, and the limit $\overline{\tau}\to\infty$ then reduces the theory to a simpler topological theory. For other recent developments in geometric quantization and its connection to topological string theory, see [66].

Even if our conjecture about the correspondence between Chern–Simons theory and the low-energy limit of the S-duality and R-symmetry twisted S^1 compactification of $\mathcal{N}=4$ SYM turns out to be wrong, it would still perhaps be interesting to explore the three-dimensional topological theory that the twisted compactification defines. There are quite a few topological quantities that can be defined through this setting. For example, we can study compactification on a Riemann surface with electric and magnetic fluxes other than selfdual combinations listed in §3.5. The mismatch between the fluxes of the original and the dual theories then needs to be corrected by inserting Wilson and/or 't Hooft line operators, as outlined at the end of §3.5. Even on $\mathbb{R}^3$, the twisted $\mathcal{N}=4$ compactification also defines expectation values for knots, which we expect to be topological, at least for the limited list of ranks and twists listed in §6.1. It would perhaps be interesting to check if these knot invariants agree with those calculated from Chern–Simons theory, or if they give rise to different knot invariants. For recent developments on the connection between knot invariants and string theory, see [67, 68]. More generally, it would perhaps be interesting to study the partition function of the twisted compactification on $M_3\times S^1$, where M_3 is a general 3-manifold. It would also be interesting to extend the discussion to selfdual theories with less supersymmetry. It was recently shown that a Chern-Simons term in three-dimensions can be induced in certain circle compactifactions of chiral four dimensional theories with a flavor symmetry twist [69]. It would perhaps be interesting to generalize this to include a duality twist.

In order to get a topological low-energy theory we had to restrict the rank of the $SU(n)$ gauge group to $n\leq 5$ (see §6.1). For higher values of n (and even for lower values, for some of the $\mathrm{SL}(2,\mathbb{Z})$ twists) we get scalar and spinor zero modes, and the low-energy description is not topological. Nonetheless, we get in this way nontrivial three-dimensional theories with $\mathcal{N}=6$ supersymmetry, and it would perhaps be interesting to

explore their connection to the superconformal theories that were recently discovered in connection with M2-branes at orbifold singularities and supersymmetrization of Chern–Simons theory [70]-[75].

More ambitiously, we would like to gain new information about S-duality itself. For this we need to understand the full topological structure of the S-duality kernel $\mathcal{S}(V, \widetilde{V})$. One possible direction might be to start with the assumption (3.5) about the expectation value of pairs of Wilson loops and attempt to reconstruct $\mathcal{S}(V, \widetilde{V})$ from it. This can be done, in principle, on a lattice, but it would be perhaps interesting to study if it has a meaningful continuum limit.

Acknowledgments
We are grateful to Mina Aganagic, Chris Beem, Aaron Bergman, Eric Gimon, Måns Henningson, Petr Hořava, Soo-Jong Rey, Mithat Ünsal, David Vegh, and Edward Witten for discussions and correspondence. OJG wishes to thank Rob Myers, Mark van Raamsdonk and Wati Taylor, the organizers of the workshop on "Emerging Directions in String Theory" (where preliminary results of this work have been presented), and the Banff International Research Station for Mathematical Innovation and Discovery (BIRS) for their hospitality. This work was supported in part by the Berkeley Center of Theoretical Physics, in part by the U.S. National Science Foundation under grant PHY-04-57315, and in part by the Director, Office of Science, Office of High Energy Physics, of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

A. Chern–Simons theory with magnetic flux

The dimension of the Hilbert space of $SU(2)$ Chern–Simons theory at level k on a Riemann surface of genus h is given by [49]:

$$d_h(2,k) = \left(\frac{k}{2}+1\right)^{h-1} \sum_{j=0}^{k} \frac{1}{\sin^{2(h-1)}(\frac{j+1}{k+2})\pi} \,. \tag{A.1}$$

This, by definition, has no magnetic flux. However, in this paper we need the dimension of the Hilbert space of $SU(2)/\mathbb{Z}_2$ gauge configurations with one unit of magnetic flux ($\mathbf{m}_0 = 1$ and $\mathbf{m}_1 = 0$ in the notation of §6.6). In this appendix we will outline the calculation of this dimension, following [25][49]. We focus on the case of genus $h = 2$, but it is easy to generalize to higher genus.

The dimension $d_h(2, k, \mathbf{m})$ is equal to the partition function on $\mathcal{C}_2 \times S^1$ with the gauge bundle specified according to the magnetic flux $\mathbf{m}$. We refer to S^1 as (Euclidean) "time." The partition function can be calculated by cutting the Riemann surface along

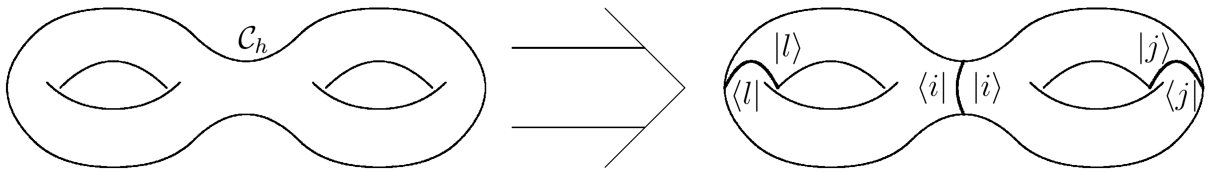

Figure 2: The partition function of Chern–Simons theory on $\mathcal{C}_2$ times S^1 (not shown) is calculated as follows [25]: (i) the Riemann surface $\mathcal{C}_2$ is cut in three places along three circles; (ii) a complete set of states (of the Hilbert space on T^2 which is the cut times the S^1 that is not shown) is inserted on the two sides of each cut, $|l\rangle\langle l|, |i\rangle\langle i|, |j\rangle\langle j|$; (iii) the cuts are separated to form two 3-holed spheres and the partition function is calculated using the fusion rules as $\sum_{ijl} N_{lli}N_{ijj}$.

three loops, as in Figure 2. Each cut has the topology of a torus $S^1 \times S^1$, where the second factor is time and the first factor is the appropriate loop that corresponds to one of the three one-cycles along which we cut. Thus $\mathcal{C}_2$ is realized as two 3-holed spheres glued appropriately along the holes. Next, for each cut (hole) we insert a complete set of states $|i\rangle\langle i|$ of the Hilbert space of Chern–Simons theory on $S^1 \times S^1$ and calculate the partition function using the known expressions for the partition function of Chern–Simons theory on a 3-holed sphere with boundary states $|i_1\rangle, |i_2\rangle, |i_3\rangle$, i.e., the "fusion rules" $N_{i_1 i_2 i_3}$. If we insert the complete sets of states $|l\rangle\langle l|, |i\rangle\langle i|, |j\rangle\langle j|$ in the left, middle, and right cut, respectively (see Figure 2), then the partition function without magnetic flux can be written as as $\sum_{ijl} N_{lli}N_{ijj}$.

To introduce magnetic flux $\mathbf{m}_0 = 1$ we can insert a large gauge transformation along one of the cuts, say the middle one, before gluing. The large gauge transformation corresponds to a topologically nontrivial map $S^1 \to SO(3)$ and if we denote its action on the states by $|i\rangle \mapsto \sum_p \Lambda_{ip}|p\rangle$ we get the partition function

$$d_2(2, k, \mathbf{m}_0 = 1) = \sum_{ipjl} N_{lli}\Lambda_{ip}N_{pjj}.$$

To get explicit expressions we use the basis of states on $T^2 \simeq S^1 \times S^1$ introduced in [25]. The states are denoted by $|l\rangle$ with $l = 0, \ldots, k$ labeling a representation of spin $l/2$. The state $|l\rangle$ can be realized by filling the first S^1 factor to form a disc, then inserting a closed Wilson line in the representation with spin $l/2$. If we assume that the Wilson line runs in the time direction (the second S^1 factor) and is located at, say, the origin

of the disc then the fusion rules have a simple expression [76]:

$$N_{i_1 i_2 i_3} = \begin{cases} 1 \text{ if } |i_2 - i_1| \le i_3 \le i_1 + i_2, \text{ and } i_1 + i_2 + i_3 \le 2k; \\ 0 \text{ otherwise.} \end{cases}$$

On the other hand, if we define the basis of states by filling the second (time direction) S^1 instead, and let the Wilson lines run parallel to the first S^1 factor, we get a basis of states $|i'\rangle$ on which the large gauge transformation is easy to describe (since it acts on the Wilson line in a simple way):

$$\Lambda_{i'p'} = \delta_{i'p'}(-1)^{i'}.$$

The unitary transformation from the basis $\{|j\rangle\}$ to the basis $\{|j'\rangle\}$ is given by [76]:

$$|j'\rangle = \sum_j S_{j'j}|j\rangle, \qquad S_{j'j} = \sin \frac{(j+1)(j'+1)\pi}{k+2}.$$

Using this expression the partition function can be calculated as

$$d_2(2, k, \mathbf{m}_0 = 1) = \sum_l (-1)^l \left(\sum_p S_{lp} \sum_j N_{pjj} \right)^2.$$

For $k = 2$ we get $d_2(2, 2, \mathbf{m}_0 = 1) = 6$. This agrees with the result of [77].

References

[1] C. Montonen and D. I. Olive, "Magnetic Monopoles As Gauge Particles?," Phys. Lett. B **72**, 117 (1977).

[2] P. Goddard, J. Nuyts and D. I. Olive, "Gauge Theories And Magnetic Charge," Nucl. Phys. B **125**, 1 (1977).

[3] H. Osborn, "Topological Charges For $N = 4$ Supersymmetric Gauge Theories And Monopoles Of Spin 1," Phys. Lett. B **83**, 321 (1979).

[4] A. Sen, "Dyon - monopole bound states, selfdual harmonic forms on the multi-monopole moduli space, and $SL(2, Z)$ invariance in string theory," Phys. Lett. B **329**, 217 (1994) [arXiv:hep-th/9402032].

[5] C. Vafa and E. Witten, "A Strong coupling test of S duality," Nucl. Phys. B **431**, 3 (1994) [arXiv:hep-th/9408074].

[6] M. Bershadsky, A. Johansen, V. Sadov and C. Vafa, "Topological reduction of 4-d SYM to 2-d sigma models," Nucl. Phys. B **448**, 166 (1995) [arXiv:hep-th/9501096].

[7] J. A. Harvey, G. W. Moore and A. Strominger, “Reducing S Duality To T Duality,” Phys. Rev. D **52**, 7161 (1995) [arXiv:hep-th/9501022].

[8] N. Seiberg and E. Witten, “Monopoles, duality and chiral symmetry breaking in $N = 2$ supersymmetric QCD,” Nucl. Phys. B **431**, 484 (1994) [arXiv:hep-th/9408099].

[9] K. A. Intriligator, “Bonus symmetries of $N = 4$ super-Yang–Mills correlation functions via AdS duality,” Nucl. Phys. B **551**, 575 (1999) [arXiv:hep-th/9811047];
K. A. Intriligator and W. Skiba, “Bonus symmetry and the operator product expansion of N = 4 super-Yang–Mills,” Nucl. Phys. B **559**, 165 (1999) [arXiv:hep-th/9905020].

[10] A. Kapustin, “Wilson-’t Hooft operators in four-dimensional gauge theories and S-duality,” Phys. Rev. D **74**, 025005 (2006) [arXiv:hep-th/0501015].

[11] M. Henningson and N. Wyllard, “Zero-energy states of $N = 4$ SYM on T^3: S-duality and the mapping class group,” JHEP **0804**, 066 (2008) [arXiv:0802.0660 [hep-th]]; “Bound states in $N = 4$ SYM on T^3: Spin$(2n)$ and the exceptional groups,” JHEP **0707**, 084 (2007) [arXiv:0706.2803 [hep-th]]; “Low-energy spectrum of $N = 4$ super-Yang–Mills on T^3: Flat connections, bound states at threshold, and S-duality,” JHEP **0706**, 001 (2007) [arXiv:hep-th/0703172].

[12] D. Gaiotto and E. Witten, “Supersymmetric Boundary Conditions in $N = 4$ Super Yang-Mills Theory,” arXiv:0804.2902 [hep-th].

[13] D. Gaiotto and E. Witten, “S-Duality of Boundary Conditions In $N = 4$ Super Yang-Mills Theory,” arXiv:0807.3720 [hep-th].

[14] O. Aharony, S. S. Gubser, J. M. Maldacena, H. Ooguri and Y. Oz, “Large N field theories, string theory and gravity,” Phys. Rept. **323**, 183 (2000) [arXiv:hep-th/9905111].

[15] E. Witten, “Some comments on string dynamics,” arXiv:hep-th/9507121.

[16] C. Vafa, “Geometric origin of Montonen-Olive duality,” Adv. Theor. Math. Phys. **1**, 158 (1998) [arXiv:hep-th/9707131].

[17] T. Banks, W. Fischler, S. H. Shenker and L. Susskind, “M theory as a matrix model: A conjecture,” Phys. Rev. D **55**, 5112 (1997) [arXiv:hep-th/9610043].

[18] O. Aharony, M. Berkooz, S. Kachru, N. Seiberg and E. Silverstein, “Matrix description of interacting theories in six dimensions,” Adv. Theor. Math. Phys. **1**, 148 (1998) [arXiv:hep-th/9707079].

[19] O. Aharony, M. Berkooz and N. Seiberg, “Light-cone description of (2,0) superconformal theories in six dimensions,” Adv. Theor. Math. Phys. **2**, 119 (1998) [arXiv:hep-th/9712117].

[20] O. J. Ganor and S. Sethi, "New perspectives on Yang–Mills theories with sixteen supersymmetries," JHEP **9801**, 007 (1998) [arXiv:hep-th/9712071].

[21] A. Kapustin and E. Witten, "Electric-magnetic duality and the geometric Langlands program," arXiv:hep-th/0604151.

[22] D. Karabali and V. P. Nair, "A gauge-invariant Hamiltonian analysis for non-Abelian gauge theories in (2+1) dimensions," Nucl. Phys. B **464**, 135 (1996) [arXiv:hep-th/9510157]; "On the origin of the mass gap for non-Abelian gauge theories in (2+1) dimensions," Phys. Lett. B **379**, 141 (1996) [arXiv:hep-th/9602155]; D. Karabali, C. j. Kim and V. P. Nair, "Planar Yang–Mills theory: Hamiltonian, regulators and mass gap," Nucl. Phys. B **524**, 661 (1998) [arXiv:hep-th/9705087]; "On the vacuum wave function and string tension of Yang–Mills theories in (2+1) dimensions," Phys. Lett. B **434**, 103 (1998) [arXiv:hep-th/9804132]; "Manifest covariance and the Hamiltonian approach to mass gap in (2+1)-dimensional Yang–Mills theory," Phys. Rev. D **64**, 025011 (2001) [arXiv:hep-th/0007188].

[23] R. G. Leigh and D. Minic, "Towards the QCD string: 2+1 dimensional Yang–Mills theory in the planar limit," arXiv:hep-th/0407051; R. G. Leigh, D. Minic and A. Yelnikov, "Solving pure QCD in 2+1 dimensions," Phys. Rev. Lett. **96**, 222001 (2006) [arXiv:hep-th/0512111]; "On the spectrum of Yang–Mills theory in 2+1 dimensions, analytically," arXiv:0704.3694 [hep-th].

[24] L. Freidel, R. G. Leigh, D. Minic and A. Yelnikov, "On the spectrum of pure Yang–Mills theory," arXiv:0801.1113 [hep-th].

[25] E. Witten, "Quantum field theory and the Jones polynomial," Commun. Math. Phys. **121**, 351 (1989).

[26] C. Vafa, "Evidence for F-Theory," Nucl. Phys. B **469**, 403 (1996) [arXiv:hep-th/9602022].

[27] A. Dabholkar and J. A. Harvey, "String islands," JHEP **9902**, 006 (1999) [arXiv:hep-th/9809122].

[28] O. J. Ganor, "U-duality twists and possible phase transitions in (2+1)D supergravity," Nucl. Phys. B **549**, 145 (1999) [arXiv:hep-th/9812024].

[29] J. Shelton, W. Taylor and B. Wecht, "Nongeometric Flux Compactifications," JHEP **0510**, 085 (2005) [arXiv:hep-th/0508133].

[30] D. Vegh and J. McGreevy, "Semi-Flatland," JHEP **0810**, 068 (2008) [arXiv:0808.1569 [hep-th]].

[31] A. Strominger, "Open p-branes," Phys. Lett. B **383**, 44 (1996) [arXiv:hep-th/9512059].

[32] E. Witten, "On the Landau-Ginzburg description of $N = 2$ minimal models," Int. J. Mod. Phys. A **9**, 4783 (1994) [arXiv:hep-th/9304026].

[33] G. 't Hooft, "On The Phase Transition Towards Permanent Quark Confinement," Nucl. Phys. B **138**, 1 (1978); "A Property Of Electric And Magnetic Flux In Nonabelian Gauge Theories," Nucl. Phys. B **153**, 141 (1979).

[34] Work in progress.

[35] J. Polchinski, *String Theory*, Vol. 1 (Cambridge University Press, 1998).

[36] A. Giveon, M. Porrati and E. Rabinovici, "Target space duality in string theory," Phys. Rept. **244**, 77 (1994) [arXiv:hep-th/9401139].

[37] E. Witten, "Mirror manifolds and topological field theory," arXiv:hep-th/9112056.

[38] Y. Lozano, "S duality in gauge theories as a canonical transformation," Phys. Lett. B **364**, 19 (1995) [arXiv:hep-th/9508021].

[39] A. Giveon and E. Witten, "Mirror symmetry as a gauge symmetry," Phys. Lett. B **332**, 44 (1994) [arXiv:hep-th/9404184].

[40] D. R. Morrison and M. R. Plesser, "Towards mirror symmetry as duality for two dimensional abelian gauge theories," Nucl. Phys. Proc. Suppl. **46**, 177 (1996) [arXiv:hep-th/9508107].

[41] A. S. Schwarz, "Lefschetz trace formula and BRST," Mod. Phys. Lett. A **4**, 1891 (1989).

[42] A. J. Niemi and K. Palo, "On quantum integrability and the Lefschetz number," Mod. Phys. Lett. A **8**, 2311 (1993) [arXiv:hep-th/9305077].

[43] S. Li, "Supersymmetric quantum mechanics and Lefschetz fixed-point formula," arXiv:hep-th/0511101.

[44] E. Witten, "On S duality in Abelian gauge theory," Selecta Math. **1**, 383 (1995) [arXiv:hep-th/9505186].

[45] E. Witten, "SL$(2, \mathbb{Z})$ action on three-dimensional conformal field theories with Abelian symmetry," arXiv:hep-th/0307041.

[46] S. J. Rey and J. T. Yee, "Macroscopic strings as heavy quarks in large N gauge theory and anti-de Sitter supergravity," Eur. Phys. J. C **22**, 379 (2001) [arXiv:hep-th/9803001].

[47] J. M. Maldacena, "Wilson loops in large N field theories," Phys. Rev. Lett. **80**, 4859 (1998) [arXiv:hep-th/9803002].

[48] M. Bershadsky, C. Vafa and V. Sadov, "D-Branes and Topological Field Theories," Nucl. Phys. B **463**, 420 (1996) [arXiv:hep-th/9511222].

[49] H. L. Verlinde and E. P. Verlinde, "Conformal field theory and geometric quantization," PUPT-89-1149, IASSNS-HEP-89-58, Oct 1989.

[50] N. J. Hitchin, "The Selfduality Equations On A Riemann Surface," Proc. Lond. Math. Soc. **55**, 59 (1987).

[51] T. Hausel and M. Thaddeus, "Generators for the cohomology ring of the moduli space of rank 2 Higgs bundles," arXiv:math/0003093; "Relations in the cohomology ring of the moduli space of rank 2 Higgs bundles" arXiv:math/0003094.

[52] P. E. Newstead, "Characteristic classes of stable bundles of rank 2 over an algebraic curve," Trans. Amer. Math. Soc. **169**, 337 (1972).

[53] E. Witten, "Two-dimensional gauge theories revisited," J. Geom. Phys. **9** (1992) 303 [arXiv:hep-th/9204083].

[54] A. D. King and P. E. Newstead, "On the cohomology ring of the moduli space of rank 2 vector bundles on a curve," Topology **37**, 407 (1998).

[55] T. Hausel, "Vanishing of intersection numbers on the moduli space of Higgs bundles," Adv. Theor. Math. Phys. **2**, 1011 (1998) [arXiv:math/9805071].

[56] T. Hausel and M. Thaddeus, "Mirror symmetry, Langlands duality, and the Hitchin system," arXiv:math/0205236.

[57] A. Strominger, S. T. Yau and E. Zaslow, "Mirror symmetry is T-duality," Nucl. Phys. B **479**, 243 (1996) [arXiv:hep-th/9606040].

[58] K. Hori and C. Vafa, "Mirror symmetry," arXiv:hep-th/0002222.

[59] M. Aganagic and C. Vafa, "Perturbative derivation of mirror symmetry," arXiv:hep-th/0209138.

[60] A. Hashimoto and K. Thomas, "Non-commutative gauge theory on D-branes in Melvin universes," JHEP **0601**, 083 (2006) [arXiv:hep-th/0511197].

[61] E. Witten, "New 'gauge' theories in six dimensions," JHEP **9801**, 001 (1998) [Adv. Theor. Math. Phys. **2**, 61 (1998)] [arXiv:hep-th/9710065].

[62] D. Gaiotto and E. Witten, "Janus Configurations, Chern-Simons Couplings, And The Theta-Angle in $N = 4$ Super Yang-Mills Theory," arXiv:0804.2907 [hep-th].

[63] T. Kitao, K. Ohta and N. Ohta, "Three-dimensional gauge dynamics from brane configurations with (p, q)-fivebrane," Nucl. Phys. B **539**, 79 (1999) [arXiv:hep-th/9808111].

[64] P. Horava, "Quantum Criticality and Yang–Mills Gauge Theory," arXiv:0811.2217 [hep-th].

[65] E. Frenkel, A. Losev and N. Nekrasov, "Notes on instantons in topological field theory and beyond," Nucl. Phys. Proc. Suppl. **171**, 215 (2007) [arXiv:hep-th/0702137]; "Instantons beyond topological theory II," arXiv:0803.3302 [hep-th].

[66] S. Gukov and E. Witten, "Branes and Quantization," arXiv:0809.0305 [hep-th].

[67] S. Gukov, A. Iqbal, C. Kozcaz and C. Vafa, "Link homologies and the refined topological vertex," arXiv:0705.1368 [hep-th].

[68] S. Gukov, "Surface Operators and Knot Homologies," arXiv:0706.2369 [hep-th].

[69] E. Poppitz and M. Ünsal, "Index theorem for topological excitations on $R^3 \times S^1$ and Chern-Simons theory," arXiv:0812.2085 [hep-th].

[70] J. H. Schwarz, "Superconformal Chern–Simons theories," JHEP **0411**, 078 (2004) [arXiv:hep-th/0411077].

[71] A. Gustavsson, "Algebraic structures on parallel M2-branes," arXiv:0709.1260 [hep-th].

[72] J. Bagger and N. Lambert, "Comments On Multiple M2-branes," JHEP **0802**, 105 (2008) [arXiv:0712.3738 [hep-th]].

[73] N. Lambert and D. Tong, "Membranes on an Orbifold," Phys. Rev. Lett. **101**, 041602 (2008) [arXiv:0804.1114 [hep-th]].

[74] J. Distler, S. Mukhi, C. Papageorgakis and M. Van Raamsdonk, "M2-branes on M-folds," JHEP **0805**, 038 (2008) [arXiv:0804.1256 [hep-th]].

[75] O. Aharony, O. Bergman, D. L. Jafferis and J. Maldacena, "N=6 superconformal Chern–Simons-matter theories, M2-branes and their gravity duals," JHEP **0810**, 091 (2008) [arXiv:0806.1218 [hep-th]].

[76] E. P. Verlinde, "Fusion Rules And Modular Transformations In 2d Conformal Field Theory," Nucl. Phys. B **300**, 360 (1988).

[77] G. Daskalopoulos and R. Wentworth, "The integrability criterion in SU(2) Chern–Simons gauge theory," Commun. Math. Phys. **151**, 437 (1993).